Identifying a Minimally Important Change Threshold for The Multiple Sclerosis Walking Scale-12 (MSWS-12)

Mehta, L

http://hdl.handle.net/10026.1/5550

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.
Identifying an Important Change Estimate for the Multiple Sclerosis Walking Scale-12 (MSWS-12v1) For Interpreting Clinical Trial Results

Lahar Mehta, MD¹; Manjit McNeill, MSc²; Jeremy Hobart, FRCP, PhD³; Kathleen Wyrwich, PhD⁴; Jiat-Ling Poon, PhD⁴; Priscilla Auguste, MHS⁴; John Zhong, PhD¹; Jacob Elkins, MD¹

¹Biogen, Cambridge, MA, USA; ²Biogen, Maidenhead, United Kingdom;
³Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK;
⁴Evidera, Bethesda, MD, USA

Running Title: Important Change Estimate for MSWS-12v1

Keywords: important change, meaningful improvement, responder definition, MSWS-12, patient-reported outcomes, multiple sclerosis, walking ability, anchor-based, distribution-based

Corresponding Author:

Kathleen W. Wyrwich, PhD
7101 Wisconsin Avenue, Suite 1400
Bethesda, MD 20814
USA
kathy.wyrwich@evidera.com
This is an author’s draft of an accepted article submitted and published in Multiple Sclerosis Journal

DOI: http://mso.sagepub.com/content/1/2055217315596993.full

+1-314-773-8000 (phone)

+1-314-773-8004 (fax)
ABSTRACT

Background: The 12-question Multiple Sclerosis (MS) Walking Scale (MSWS-12v1) is a widely-used patient-reported outcome (PRO) measure of walking ability in MS.

Objective: To estimate the magnitude of an important change in MSWS-12v1 scores for the interpretation of meaningful subject-level improvements across a 6-month trial of MS patients with walking disability.

Methods: MOBILE was a 6-month exploratory study assessing fampridine’s effect on walking ability in 132 people with MS. Three PRO measures assessed walking ability: MSWS-12v1, EuroQol 5-Dimension-5 Level (EQ-5D-5L) mobility question, and a patient global impression of change (PGIC) in overall walking ability. Pre-specified anchor- and distribution-based analyses estimated the MSWS-12v1 change scores representing an important change for participants. Results were triangulated to propose a single best value indicating meaningful improvement.

Results: Using Baseline to Week 2 through Week 24 change scores, anchor-based analyses demonstrated mean and median improvements of 5.2-6.6 (PGIC) and 9.7-13.4 (EQ-5D-5L mobility) points on the MSWS-12v1 indicated meaningful improvements. The distribution-based estimate was 6.8 points. Triangulation across the results suggested an 8-point reduction in MSWS-12v1 score represents an important subject-level change in these participants.

Conclusion: In similar MS clinical trials, an 8-point improvement on the MSWS-12v1 is a reasonable estimate of meaningful improvement in walking ability.
INTRODUCTION

Patient-reported outcomes (PROs) are increasingly used as efficacy endpoints in clinical trials to measure symptoms and disease impact. It is therefore useful to understand the clinical importance or meaningfulness of changes in PRO scores to the individuals receiving treatments. The concept of a minimal clinically important difference (MCID) was suggested to address the need for meaningful interpretation of PRO scores. Initially, the MCID was defined as the ‘‘smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management’’ 1. More recently, a similar concept has been referred to as the responder definition (RD), or “the individual patient PRO score change over a predetermined time period that should be interpreted as a treatment benefit” 2. Both of these terms seek to understand the magnitude of a PRO measure’s change over time that demonstrates a meaningful improvement to an individual patient.

The widely-used 12-question Multiple Sclerosis (MS) Walking Scale (MSWS-12v1) is a patient-reported measure of walking ability in MS 3. Evidence from multiple studies supports its robust measurement performance 3-8. With approximately 75% of people with MS identifying gait, mobility, and balance as key physical problems, the MSWS-12v1 has been used frequently in contemporary MS clinical trials to assess these important issues related to walking ability 9-12. The objective of this current analysis was to estimate the MSWS-12v1 change score representing a meaningful improvement in walking ability, using standard methods, in data from walking-disabled people with MS participating in a 6-month, randomized treatment trial.
METHODS

Participants
Data from MOBILE (NCT01597297; EudraCT 2012-000368-90) are reported elsewhere \(^{13}\). Briefly, it was a multicenter, randomized, double-blind, placebo-controlled trial to assess the longer-term effects of fampridine-PR 10 mg BD tablets on self-assessed walking disability, and the impact of treatment on overall walking ability in MS over 24 weeks. A total of 132 walking-disabled people with MS, aged 18 to 70, were enrolled at 24 sites in Belgium, Canada, Italy, the Netherlands, Sweden, and the United Kingdom. Eligibility criteria were: any MS subtype, screening visit Expanded Disability Status Scale (EDSS) \(^ {14}\) score of 4 to 7 and clinically stable (no MS exacerbation within 60 days of screening). Participants were randomized in a 1:1 ratio on Day 1 to receive either fampridine-PR 10 mg BD or matching placebo. As shown in Figure 1, scheduled study visits were at Day 1, Weeks 2, 4, 8, 12, 16, 20, 24, and 26 (two weeks post treatment discontinuation). Unscheduled visits (if required) occurred within five days of the onset of symptoms indicating possible relapse, suspected seizure, or possible deterioration of renal function.

Measures

Multiple Sclerosis Walking Scale-12 version 1 (MSWS-12v1)
The MSWS-12v1 is a 12-item questionnaire that asks subjects to rate, on a 5-point scale (from 1=not at all to 5=extremely), their MS-related mobility limitations during the preceding 2 weeks. The MSWS-12v1 was completed at Screening and all scheduled and unscheduled study visits.

To calculate the MSWS-12v1 score, the sum of the 12 questions (sum range: 12-60) is
transformed to have a range of 0 to 100, where higher scores indicate greater walking limitations. For individuals completing all 12 items at each study visit, the possible score changes over time increase or decrease in increments of ~2 points (also known as a PRO state change) \(^{15}\) on the 0-100 point scale. For visits where responses to \(\leq 6\) of 12 MSWS-12v1 questions were missing, the person-specific mean score from the answered questions was imputed as the score for each missing question. For visits where \(\geq 7\) of the 12 component questions were not answered, the MSWS-12v1 score was considered missing \(^3\).

The MSWS-12v1 includes an initial screening question asking subjects if they are unable to walk. People responding “Yes” were instructed not to complete the questionnaire. Their MSWS-12v1 score was set to 100 if none of the MSWS-12v1 questions were completed; however, few subjects (n =3) replied in this manner during the study. Irrespective of the response to this initial question, if subjects responded to the 12 questions, these were used in deriving the MSWS-12v1 score.

**Patient Global Impression of Change (PGIC) in Walking**

A Patient Global Impression of Change (PGIC) assessment of walking ability is a single question with multiple response options. The PGIC used in the MOBILE trial asked: “In the past 7 days, how much has the study drug affected your overall walking?” Response options were: 1=very much worse, 2=much worse, 3=slightly worse, 4=unchanged, 5=slightly improved, 6=much improved, 7=very much improved. Participants completed the PGIC at Weeks 2, 4, 8, 12, 16, 20, and 24.
EuroQoL 5-Dimension 5-Level (EQ-5D-5L)
The EQ-5D-5L is a generic self-reported measure of health status. It includes five questions concerning mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. A summary utility index value can be computed from subject’s response to these five questions; values range from -0.594 (worst possible health score) to 0 (death) to 1.000 (best health). Due to the direct relevance of the response options to walking ability, only the single EQ-5D-5L domain question measuring mobility (five-level response options: 1=no problems in walking about, 2=slight problems in walking about, 3=moderate problems in walking about, 4=severe problems in walking about, or 5=unable to walk about) was used as an anchor to estimate an important change in MSWS-12v1 score. The EQ-5D-5L was administered on Day 1, and at Weeks 4, 8, 12, 16, 20, and 24.

Multiple Sclerosis Impact Scale version 1 (MSIS-29v1)
The MSIS-29v1 is a 29-item questionnaire measuring the physical (20 items) and psychological (9 items) impact of MS. Each item has 5 response options scored from 1=not at all limited/bothered to 5=extremely limited/bothered. MSIS-29v1 subscale scores (physical or psychological) are calculated by transforming the sum of the subscale items to a score ranging from 0 (no impact of MS) to 100 (extreme impact of MS). The MSIS-29v1 was completed at Screening, and at each scheduled and unscheduled visit.

Statistical Analyses
All analyses were pre-specified. Both anchor- and distribution-based methods were used to inform the important change estimate for the MSWS-12v1 score associated with a meaningful
improvement in walking ability during the MOBILE study. Negative MSWS-12v1 change scores indicate improvements in walking.

**Anchor-based Analyses**

MSWS-12v1 change scores were correlated with the proposed anchors that directly assessed walking ability (PGIC scores and EQ-5D-5L mobility item change scores) to determine if they met the suggestion that associations should exceed 0.30-0.35 \(^{18}\).

**PGIC**

The PGIC was used as one anchor to estimate the MSWS-12v1 score change associated with a meaningful patient-reported change in walking. The retrospective PGIC responses of ‘5=slightly improved’ were selected to represent time points where participants registered meaningful changes in walking ability. Each time a participant gave a PGIC score of ‘5=slightly improved’ we computed the MSWS-12v1 change score from the preceding visit. As the PGIC was collected at seven visits, each participant could contribute between zero and seven MSWS-12v1 change scores. To ensure that each participant contributed only one MSWS-12v1 change score to the group-level analyses we generated a single value for participants who gave at least one PGIC responses of ‘5=slightly improved’. Specifically, we computed the mean (Method 1) and median (Method 2) of the individual’s MSWS-12v1 change scores. Group level MSWS-12v1 change scores were then computed (Method 1 = mean [SD], range; Method 2 = median [IQR], range).

**EQ-5D-5L Mobility Question**

As a second anchor-based approach, the MSWS-12v2 change scores associated with meaningful changes in cross-sectional EQ-5D-5L mobility question were examined. The EQ-5D-5L was
collected at seven time points, giving each participant up to six EQ-5D-5L change scores, computed from baseline (e.g., EQ-5D-5L score at Week 4 minus EQ-5D-5L score at Day 1 yielded one EQ-5D-5L change score). For each participant, the median EQ-5D-5L change score across visits was computed. Participants with a median EQ-5D-5L change score = 1 were considered to have reported an important change. For these participants, the mean MSWS-12v1 change scores across visits were determined. Finally, across participants with median EQ-5D-5L change scores = 1 (one level of improvement on the mobility scale), mean and median of the mean MSWS-12v1 change scores were computed.

**Distribution-based Approach**

The standard error of measurement (SEM) was calculated as the distribution-based estimate of meaningful change for the MSWS-12v1. Using data from Screening and Day 1, the SEM was calculated as \( \text{SEM} = \sigma_x \sqrt{1 - r_{xx}} \), where \( \sigma_x \) is the MSWS-12v1 baseline score standard deviation (Baseline = mean of Screening and Day 1 scores), and \( r_{xx} \) is the MSWS-12v1 reliability. Test-retest reliability was estimated via the intra-class correlation coefficient [ICC; method (2,1)] of the MSWS-12v1 between Screening and Day 1 for subjects who remained stable between these two time points. Stable subjects were identified using changes on the MSIS-29v1 physical subscale (\(-7.5 < \text{MSIS-29v1 physical subscale change} < 7.5\)\(^{22,23}\). In the context of MS and walking ability, the Screening to Day 1 test-retest time period of \( \leq 14 \) days is reasonable for assessing test-retest reliability in walking ability in MS\(^{24}\).
Triangulation

A single overall estimate of a meaningful change in MSWS-12v1 score was achieved by triangulating findings of the anchor- and distribution-based analyses. Triangulation had three stages. First, we examined the range of anchor- and distribution-based estimates. Second, we examined all MSWS-12v1 change scores for all individual participants at all visits from Week 2 to Week 24. Third, we looked for gaps (spaces) in the observed change scores in the range identified in stage one.

This process helps to identify potential change thresholds. For example, there could be naturally occurring change thresholds in the data, such as areas in the change score range where there are no change scores. There is also the possibility of naturally-occurring splits that create candidates for cut-point.

All analyses were pre-specified and performed using SAS® v9.3.

RESULTS

Table 1 shows the MOBILE sample baseline characteristics. There were 132 participants randomized to fampridine (n=68, 52%) or placebo (n=64; 48%). The mean age was 49.8 years (SD=9.0), with 54% female. At Baseline, the mean MSWS-12v1 score across all patients was 73.7, with a slightly lower mean score in the treatment group vs. placebo (71.7 vs. 75.9), with similar standard deviations. The baseline EQ-5D-5L scores demonstrated that both the baseline mean and median perceived health status were very near the midpoint between 1.00 (full health)
and 0 (death). Only 8 subjects (6%) had any missing data on the MSWS-12v1, each with only one missing item, and hence, limiting the impact of data imputations.

**Anchor-based Approach**

Correlations between MSWS-12v1 change scores and proposed anchors (PGIC, r=-0.45; EQ-5D-5L mobility question, r=-0.42) exceeded the suggested level of 0.30-0.35\(^{18}\). This provided support for using these anchors to estimate meaningful MSWS-12v1 change scores.

Table 2 displays the MSWS-12v1 change scores (mean, SD, median, and inter-quartile ranges [IQR]) associated with each PGIC category. The number of subjects with at least one response in their Week 2 through Week 24 data in each of these PGIC response categories is provided in the second column of the table. Subjects with more than one response in any PGIC category had a single MSWS-12v1 change score data point computed for that category: either the mean (Method 1) or the median (Method 2) of their MSWS-12v1 change scores from the prior visit for the PGIC category (Table 2).

Eighty-one subjects reported, at least once, that they had ‘slightly improved’ since the previous study visit (51 subjects had two or more ‘slightly improved’ ratings). Using Method 1, the associated mean and median change (reduction) on the MSWS-12v1 for these ‘slightly improved’ subjects was 6.6 and 6.3 points. Under Method 2, the mean and median MSWS-12v1 change scores were 6.4 and 5.2 point reductions, respectively.

The second anchor was change on the EQ-5D-5L mobility question. Twenty-one subjects had a median 1-level improvement on the EQ-5D-5L mobility question over all study visits where both
EQ-5D-5L and MSWS-12v1 data were collected. Mean and median change scores for this improved group were -13.4 and -9.7, respectively. Due to the small number of individuals with this level of improvement, the median MSWS-12v1 change of -9.7 points was deemed the most appropriate meaningful change estimate using this anchor.

**Distribution-based Approach**
The standard deviation for the MSWS-12v1 at Baseline (19.6) and reliability calculated using MSWS-12v1 test-retest data from stable subjects (n=62; ICC=0.88) were used to calculate the SEM of the MSWS-12v1. The SEM estimate for the MSWS-12v1 was 6.8 points.

**Triangulation**
In order to propose a single estimate to represent a clinically meaningful improvement in MSWS-12v1 score, the values obtained from both anchor-based and distribution-based approaches were compared (Table 3). In addition, individual change scores in the MOBILE study across Week 2 through Week 24 were examined for notable gaps in this relevant range (-5.2 to -9.7); the largest of the observed gaps was between the change score values of -7.9 and -8.8 points. Based on this triangulation process\(^{18,25}\), an 8-point reduction, expressed as an integer, was selected as a reasonable estimate of a meaningful subject-level improvement on the MSWS-12v1 for this sample.

**DISCUSSION**
The goal of this study was to identify a meaningful change estimate for the MSWS-12v1 to guide the interpretation of subject-level changes in MS patients with walking disability. Based on these
results, that incorporated the complementary use of both anchor- and distribution-based methods of estimation, 8 points was estimated to be a meaningful individual improvement for the MSWS-12v1 over a 24 week period. This study was unique in that important change from Baseline was investigated using PRO anchors and associated MSWS-12v1 change scores across all visits (Week 2 through Week 24) where a relevant change could be demonstrated.

It is important to note that this estimate of 8-points change can be used for determining whether individuals with MS might be considered as having experienced a meaningful change in the same context of use (target population, clinical trial design, etc.). Applying this estimate then allows a comparison of the proportions of people achieving this threshold in different study arms. It is just as important to note that the 8-point change estimate should not be interpreted as the criterion for meaningful treatment difference for group mean change comparisons. Indeed, meaningful group mean change differences are often smaller than the meaningful individual person differences. Finally, this 8-point estimate of meaningful individual response differs from the definition of responders used to identify timed walk responders in the fampridine studies.

Previous studies have examined important changes thresholds for the MSWS-12v1. One report investigated, post hoc, whether the difference of 6.9 MSWS-12v1 points, between timed-walk responders and non-responders observed in the pivotal fampridine trials, satisfied criteria as clinically significant. Data from the two phase three trials, and multiple other studies, were examined using multiple anchor and distribution-based methods. Estimates of meaningful
changes varied from as little as 4 points (using subject-reported change anchors) to as much as 21 points (comparing MSWS-12v1 mean score differences of subjects who reported that they walked unaided to subject who reported that they walked with aids). Results were triangulated and suggested that individual-person change scores of <4 points were not significant, > 6 points met criteria as significant, and 4-6 points might be considered as borderline. These individual-person level estimates were then used to assist the interpretation that the 6.9 point mean group change was meaningful. Others have inferred that this report ascertained MCID values of 4-6 points for MSWS-12v1.

In a recent study, Baert et al. investigated clinically important improvements on five walking measures, including the MSWS-12v1, in 290 people with MS from 17 European rehabilitation centers. The estimation methods included: anchor-based approaches incorporating a PGIC and a therapist’s global rating of change scale (GRS); and distribution-based analyses of two responsiveness indices, the smallest real change (SRC) parameter calculated at the individual (SRC\text{ind}) and group (SRC\text{group}) level. The MSWS-12v1 clinically important improvement estimates ranged from 10.4 to 14.1 points, and varied across subgroups.

The 8-point result in the MOBILE study is focused on a specific clinical trial population, and multiple study visits where a fast-acting medication may demonstrate a notable treatment effect at the subject level in walking ability. The contrast between this estimate of meaningful change and the Baert et al. results may be due to several factors, including: 1) differing clinical enrollment characteristics (e.g., EDSS ≤ 6.5 vs. EDSS between 4 and 7), study design, and
setting (rehab interventional study vs. Phase 2 clinical trial); and 2) key differences in methodological approaches, including: the problematic use of regression analyses\textsuperscript{30} to examine MSWS-12v1 change at specific PGIC and GRS levels; and the use of responsiveness indices\textsuperscript{31} that incorporated a statistically higher magnitude of change than the SEM\textsuperscript{21}.

This investigation of the MSWS-12v1 meaningful change scores using the MOBILE data has several valuable features. Both anchors used in this study were PROs assessing walking ability; incorporating other patient-reported, clinical or performance-based outcomes as anchors may provide different results. From a clinical relevance perspective, this study’s specific PRO anchors (PGIC and EQ-5D-5L mobility question) are directly interpretable for understanding MSWS-12v1 changes as they directly assess the impact of walking disability from the patient’s perspective.

This study has a number of limitations. The sample from which the EQ-5D-based meaningful change estimates were derived was small. The two walking ability anchors generated different estimates for meaningful change, implying that the choice of anchor influences the meaningful change estimate. Both of these findings are to be expected and served to highlight complexities associated with estimating meaningful changes.

In conclusion, this study used established anchor- and distribution-based methods to derive an estimate of meaningful change for the MSWS-12v1 using data obtained from a MS clinical trial. Results suggest that an 8-point improvement in MSWS-12v1 is a reasonable estimate of an individual-person meaningful improvement in walking disability. Further studies should
investigate this estimate’s reproducibility and applicability to aid the interpretation of other MS clinical trials.
CONFLICT OF INTEREST STATEMENT

Drs. Mehta, Elkins, and Zhong, and Ms. McNeill have received personal compensation for activities with Biogen as employees, and hold stock and/or stock options in Biogen. Dr. Hobart has consulted for Biogen and received support for travel, research, and clinical service development from Biogen. Drs. Wyrwich and Poon, and Ms. Auguste have received research support from Biogen.
This is an author’s draft of an accepted article submitted and published in Multiple Sclerosis Journal

DOI: http://mso.sagepub.com/content/1/2055217315596993.full

FUNDING ACKNOWLEDGEMENT

This work was supported by funding from Biogen.
REFERENCES


27. Hobart J. Prolonged-release fampridine for multiple sclerosis: was the effect on walking ability clinically significant? *26th Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS) & 15th Annual Conference of Rehabilitation in MS (RIMS)*. Gothenburg, Sweden 2010.


Table 1. Baseline Characteristics of Study Population

<table>
<thead>
<tr>
<th></th>
<th>Placebo (N=64)</th>
<th>Fampridine-PR 10 mg BID (N=68)</th>
<th>Total (N=132)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean age, years (SD)</strong></td>
<td>49.8 (9.3)</td>
<td>49.8 (8.7)</td>
<td>49.8 (9.0)</td>
</tr>
<tr>
<td><strong>Male, N (%)</strong></td>
<td>31 (48)</td>
<td>30 (44)</td>
<td>61 (46)</td>
</tr>
<tr>
<td><strong>Race, N (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>63 (98)</td>
<td>66 (97)</td>
<td>129 (98)</td>
</tr>
<tr>
<td>Asian</td>
<td>0</td>
<td>2 (3)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (2)</td>
<td>0</td>
<td>1 (&lt;1)</td>
</tr>
<tr>
<td><strong>BMI, kg/m²: Mean (SD)</strong></td>
<td>26.5 (6.2)</td>
<td>26.8 (4.9)</td>
<td>26.6 (5.6)</td>
</tr>
<tr>
<td><strong>Expanded Disability Status Scale (EDSS): Mean (SD)</strong></td>
<td>5.8 (0.9)</td>
<td>5.6 (0.9)</td>
<td>5.7 (0.9)</td>
</tr>
<tr>
<td><strong>MSWS-12v1 Baseline score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>75.9 (19.8)</td>
<td>71.7 (19.3)</td>
<td>73.7 (19.6)</td>
</tr>
<tr>
<td>Median (Q1, Q3)</td>
<td>81.3 (65.6, 90.6)</td>
<td>75.0 (64.1, 84.9)</td>
<td>78.6 (64.6, 88.0)</td>
</tr>
<tr>
<td>Min, Max</td>
<td>(8.3, 100)</td>
<td>(25.0, 100)</td>
<td>(8.3, 100)</td>
</tr>
<tr>
<td><strong>MSIS-29v1 Physical Subscale Baseline score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>53.0 (19.1)</td>
<td>50.9 (19.4)</td>
<td>51.9 (19.2)</td>
</tr>
<tr>
<td>Median (Q1, Q3)</td>
<td>57.5 (40.3, 65.6)</td>
<td>50.0 (38.1, 67.2)</td>
<td>52.5 (38.8, 66.3)</td>
</tr>
<tr>
<td>Min, Max</td>
<td>(13.1, 91.9)</td>
<td>(8.1, 100)</td>
<td>(8.1, 100)</td>
</tr>
<tr>
<td><strong>MSIS-29v1 Psychological Subscale Baseline score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>36.3 (20.0)</td>
<td>36.0 (22.2)</td>
<td>36.2 (21.1)</td>
</tr>
<tr>
<td>Median (Q1, Q3)</td>
<td>34.0 (22.9, 47.2)</td>
<td>32.6 (18.1, 50.7)</td>
<td>33.3 (20.8, 49.3)</td>
</tr>
<tr>
<td>Min, Max</td>
<td>(0.0, 93.1)</td>
<td>(1.4, 90.3)</td>
<td>(0.0, 93.1)</td>
</tr>
<tr>
<td><strong>EQ-5D-5L Baseline (Day 1) utility score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>0.51 (0.23)</td>
<td>0.54 (0.20)</td>
<td>0.52 (0.21)</td>
</tr>
<tr>
<td>Median (Q1, Q3)</td>
<td>0.55 (0.39, 0.69)</td>
<td>0.58 (0.46, 0.70)</td>
<td>0.57 (0.42, 0.69)</td>
</tr>
<tr>
<td>Min, Max</td>
<td>(-0.19, 1.00)</td>
<td>(0.04, 0.85)</td>
<td>(-0.19, 1)</td>
</tr>
</tbody>
</table>
This is an author’s draft of an accepted article submitted and published in Multiple Sclerosis Journal

DOI: http://mso.sagepub.com/content/1/2055217315596993.full

* Baseline scores calculated as the mean of Screening and Day 1 scores
BID: twice daily; BMI: Body mass index
Table 2. Summary of MSWS-12v1 Changes Scores* Across PGIC categories

Method 1: Analysis included each subject’s mean MSWS-12v1 change score from prior visit for all visits within the PGIC category

<table>
<thead>
<tr>
<th>PGIC category</th>
<th>N</th>
<th>Range</th>
<th>Mean (SD)</th>
<th>Median (Inter-Quartile Range: Q1, Q3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = very much worse</td>
<td>4</td>
<td>1.0, 10.4</td>
<td>5.7 (4.4)</td>
<td>5.7 (2.1, 9.4)</td>
</tr>
<tr>
<td>2 = much worse</td>
<td>18</td>
<td>-4.2, 31.4</td>
<td>10.6 (10.1)</td>
<td>9.4 (1.6, 16.7)</td>
</tr>
<tr>
<td>3 = slightly worse</td>
<td>46</td>
<td>-22.9, 24.0</td>
<td>4.4 (8.7)</td>
<td>4.2 (0.0, 9.9)</td>
</tr>
<tr>
<td>4 = unchanged</td>
<td>120</td>
<td>-16.7, 22.2</td>
<td>0.5 (7.0)</td>
<td>0.0 (-2.7, 4.2)</td>
</tr>
<tr>
<td>5 = slight improvement</td>
<td>81</td>
<td>-35.4, 16.7</td>
<td>-6.6 (8.2)</td>
<td>-6.3 (-12.5, -2.1)</td>
</tr>
<tr>
<td>6 = much improved</td>
<td>20</td>
<td>-29.2, 11.5</td>
<td>-6.8 (12.2)</td>
<td>-4.8 (-13.0, 2.1)</td>
</tr>
<tr>
<td>7 = very much improved</td>
<td>3</td>
<td>-8.3, 1.0</td>
<td>-4.5 (4.9)</td>
<td>-6.3 (-8.3, 1.0)</td>
</tr>
</tbody>
</table>

Method 2: Analysis included each subject’s median MSWS-12v1 change score from prior visit for all visits within the PGIC category

<table>
<thead>
<tr>
<th>PGIC category</th>
<th>N</th>
<th>Range</th>
<th>Mean (SD)</th>
<th>Median (Inter-Quartile Range: Q1, Q3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = very much worse</td>
<td>4</td>
<td>0.0, 10.4</td>
<td>5.5 (4.8)</td>
<td>5.7 (1.6, 9.4)</td>
</tr>
<tr>
<td>2 = much worse</td>
<td>18</td>
<td>-4.2, 31.4</td>
<td>10.5 (10.0)</td>
<td>9.4 (2.1, 16.7)</td>
</tr>
<tr>
<td>3 = slightly worse</td>
<td>46</td>
<td>-22.9, 24.0</td>
<td>4.6 (9.0)</td>
<td>4.2 (0.0, 11.5)</td>
</tr>
<tr>
<td>4 = unchanged</td>
<td>120</td>
<td>-18.8, 27.1</td>
<td>0.3 (7.3)</td>
<td>0.0 (-2.6, 3.7)</td>
</tr>
<tr>
<td>5 = slight improvement</td>
<td>81</td>
<td>-35.4, 16.7</td>
<td>-6.4 (8.8)</td>
<td>-5.2 (-10.4, -2.1)</td>
</tr>
<tr>
<td>6 = much improved</td>
<td>20</td>
<td>-29.2, 11.5</td>
<td>-6.0 (12.4)</td>
<td>0.0 (-13.0, 2.1)</td>
</tr>
<tr>
<td>7 = very much improved</td>
<td>3</td>
<td>-8.3, 1.0</td>
<td>-4.5 (4.9)</td>
<td>-6.3 (-8.3, 1.0)</td>
</tr>
</tbody>
</table>

* A single observation per subject is used within each PGIC category
Table 3. MSWS-12v1 Responder Definition Estimates Summary

<table>
<thead>
<tr>
<th>Measures</th>
<th>Important Change Level</th>
<th>Associated MSWS-12v1 Change Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Number of subjects contributing change scores or data)</td>
<td></td>
</tr>
<tr>
<td>PGIC:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anchor-based</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method 1</td>
<td>Slightly Improved (81)</td>
<td>-6.6 and -6.3 (mean and median)</td>
</tr>
<tr>
<td>Method 2</td>
<td></td>
<td>-6.4 and -5.2 (mean and median)</td>
</tr>
<tr>
<td>EQ-5D-5L Mobility Question</td>
<td>1 Point Median Improvement (21)</td>
<td>-13.4 and -9.7 (mean and median)</td>
</tr>
<tr>
<td>Distribution-based</td>
<td>Standard Error of Measurement (SEM) (132)</td>
<td>-6.8</td>
</tr>
</tbody>
</table>
EDSS: Expanded Disability Status Scale
MSWS-12v1 was administered at Screening, Day 1, Weeks 2, 4, 8, 12, 16, 20, 24, 26, and each unscheduled visit
PGIC was administered at Weeks 2, 4, 8, 12, 16, 20, and 24
EQ-5D-5L was administered at Day 1, Weeks 4, 8, 12, 16, 20, and 24
MSIS-29v1 was administered at Screening, Day 1, Weeks 2, 4, 8, 12, 16, 20, 24, and each unscheduled visit