Research Paper

Responder definition of the Multiple Sclerosis Impact Scale physical impact subscale for patients with physical worsening

Glenn A Phillips, Kathleen W Wyrwich, Shien Guo, Rossella Medori, Arman Altincatal, Linda Wagner and Jacob Elkins

Abstract

Background: The 29-item Multiple Sclerosis Impact Scale (MSIS-29) was developed to examine the impact of multiple sclerosis (MS) on physical and psychological functioning from a patient's perspective.

Objective: To determine the responder definition (RD) of the MSIS-29 physical impact subscale (PHYS) in a group of patients with relapsing–remitting MS (RRMS) participating in a clinical trial.

Methods: Data from the SELECT trial comparing daclizumab high-yield process with placebo in patients with RRMS were used. Physical function was evaluated in SELECT using three patient-reported outcomes measures and the Expanded Disability Status Scale (EDSS). Anchor- and distribution-based methods were used to identify an RD for the MSIS-29.

Results: Results across the anchor-based approach suggested MSIS-29 PHYS RD values of 6.91 (mean), 7.14 (median) and 7.50 (mode). Distribution-based RD estimates ranged from 6.24 to 10.40. An RD of 7.50 was selected as the most appropriate threshold for physical worsening based on corresponding changes in the EDSS (primary anchor of interest).

Conclusion: These findings indicate that a ≥7.50 point worsening on the MSIS-29 PHYS is a reasonable and practical threshold for identifying patients with RRMS who have experienced a clinically significant change in the physical impact of MS.

Keywords: Responder definition, MSIS-29, SELECT, patient-reported outcomes, physical worsening, anchor-based method, multiple sclerosis, distribution-based method

Correspondence to:
Glenn A Phillips
Health Economics and
Outcomes Research, Global
Market Access, Biogen
Idec, 133 Boston Post Road,
Weston, MA 02493, USA.
**glenn.phillips@biogenidec.
com**

Kathleen W Wyrwich
Evidera, Bethesda, MD, USA

Shien Guo
Evidera, Lexington, MA,
USA

Rossella Medori
Jacob Elkins
Biogen Idec, Cambridge,
MA, USA

Arman Altincatal
Evidera, Lexington, MA,
USA

Linda Wagner
Envision Pharma Group,
Southport, CT, USA

Introduction

Patient-reported outcomes (PROs) are often included as efficacy endpoints in clinical trials to provide insight into patients' perspectives of the impact of a disease or symptom on their life experience.[1] When using PRO data to evaluate an intervention in a clinical trial population, a change in the PRO that can identify a meaningful change to the individual must be defined to assist in the interpretation of the intervention's results. Although statistical significance in PRO change scores may be reported to compare differences between treatment groups, these results may be driven by differing sample sizes and variability. Understanding which patients achieve a change in PRO score that represents an important and non-trivial improvement or decline from their perspective is essential.

Historically, the term minimally important difference (MID) was used to refer to a change in PRO score that could be interpreted as clinically meaningful in clinical trials when comparing mean differences in active treatment with placebo.[2] However, MID also has been used to name the threshold that identifies an important level of individual change over time. In response to this dual interpretation of MID, the US Food and Drug Administration released a guidance document in 2009 for the use and interpretation of PRO scores in product development.[1] In this document, the term MID was replaced with responder definition (RD),

which was defined as 'the individual patient PRO score change over a predetermined time period that should be interpreted as a treatment benefit' as the key PRO interpretation threshold.[1]

The RD for a PRO is determined based on empirical evidence gained from employing anchor-based methodology.[3] This approach relates changes in scores for the PRO of interest to a known established magnitude of change that has clinical importance in a second measure associated with the PRO being evaluated and that is more easily interpreted than the PRO. Anchors can be patient global assessments of change, direct clinical anchors or clinician ratings, if appropriate. Use of a distribution-based approach is considered to have a supportive role in establishing an RD. In this approach, statistical parameters from the clinical trial population, standard error of measurement (SEM) and effect sizes (ES) are used to estimate clinically significant change in PRO scores.[3]

The 29-item Multiple Sclerosis Impact Scale (MSIS-29) is a disease-specific PRO measure that examines the impact of multiple sclerosis (MS) on physical and psychological functioning.[4] It is composed of two scales: the physical impact subscale (PHYS) and the psychological impact subscale.[4] Studies have shown the validity, reliability, and psychometric properties of the MSIS-29 and its relationships to other measures.[4–11] The MSIS-29 PHYS score, but not the psychological impact subscale score, correlated well with other measures of disability in MS, including the Expanded Disability Status Scale (EDSS)[5,8,9] and the MS Functional Composite.[8,12] The MSIS-29 was more responsive to change than the EDSS,[13] Short-Form Health Survey-36®,[14] and the Functional Assessment of MS instrument.[7,11,15] These findings suggest that the MSIS-29 PHYS may be useful to measure change in physical function in MS clinical trials.

Establishing an RD for the MSIS-29 PHYS would be useful for measuring change in the clinical trial setting of patients with remitting-relapsing MS (RRMS). An RD for the MSIS-29 PHYS was established in a community-based study of patients with MS across a broad range of disability using the anchor-based approach with a single disease-specific measure, the EDSS.[16] The current analysis was designed to evaluate the RD of the MSIS-29 PHYS using the anchor-based approach supported by distribution-based estimates in a more homogenous group of patients with RRMS participating in a clinical trial.

Materials and methods

Data source

Data from SELECT (NCT00390221), a randomized double-blind placebo-controlled study of daclizumab high-yield process (DAC HYP) 150 mg and 300 mg administered subcutaneously every 4 weeks in patients with RRMS, were used for this analysis.[17] Because clinical outcomes (annualized relapse rate, MRI outcomes) with DAC HYP 150 mg and 300 mg were similar, only the 150-mg dose is being studied in ongoing registration trials in MS. SELECT included multiple potential PRO anchors, including the MSIS-29, Short-Form Health Survey-12® (SF-12), and the EuroQol 5-Dimensions (EQ-5D), as well as a clinician assessment, the EDSS. The MSIS-29, SF-12, and EQ-5D were assessed at baseline and 12, 24, and 52 weeks. EDSS was assessed at baseline and 12, 20, 24, 36, 48, and 52 weeks, in addition to immediately following relapse, as an indicator of disability progression. This analysis used the modified intention-to-treat (ITT) population: all ITT patients from SELECT who received ≥1 dose of DAC HYP or placebo and completed ≥1 post-baseline (week 12, 24, or 52) MSIS-29 PHYS assessment.

The EDSS is a neurological function rating scale for MS[18] and is one of the most commonly used tools to assess physical disability in patients with MS.[19] The EDSS is administered by a trained examiner, typically a neurologist, based on a standard neurologic exam that assesses seven functional systems (pyramidal, cerebellar, brain stem, sensory, bowel and bladder, visual, and cerebral). It is scored on an ordinal scale ranging from 0 (normal neurological exam) to 10.0 (death due to MS) in 0.5-point increments. In SELECT, a three-month confirmed disability progression was the identified threshold for EDSS sustained disability progression and was defined as a ≥1.0-point increase in EDSS score for patients with a baseline EDSS score of ≥1.0, or a ≥1.5-point increase for patients with a baseline EDSS score of 0 sustained for 12 weeks.[17] Sensitivity analyses examined MSIS-29 PHYS change scores for patients with an EDSS change equal to 1.0 or 1.5 and a baseline EDSS score of 0. Confirmation of three-month disability progression had to occur >12 weeks later, at a visit when a relapse was not occurring.

The MSIS-29 PHYS is a 20-item subscale measuring physical impact of MS.[4] Items on the MSIS-29 PHYS have a Likert scale format (range 1.00–5.00); higher scores indicate a greater degree of disability. Total score is derived by summing items and transforming

them into a score out of 100; higher scores imply a greater degree of disability. At the individual level, change in the score of one response option (e.g. from 3.00 to 4.00) on one item shifts the entire score by 1.25 points; change scores on this instrument are in increments of 1.25. It is important to note that 1.25 is the minimum change possible on a completed MSIS-29 PHYS, and therefore any RD must be a multiple of 1.25.

The SF-12 is a self-administered, 12-item survey that measures general health status in eight domains (physical functioning, role functioning–physical, bodily pain, general health, vitality, social functioning, role functioning–emotional, and mental health).[20] Results of the SF-12 are expressed in terms of two summary scores: the physical component summary (PCS) and the mental component summary. SF-12 was designed to have a mean score of 50 and a standard deviation (SD) of 10 in a representative sample of the US population. Changes in score less than –3, –5, and –6 at a given time point from baseline are all possible representations of the threshold for increasing disability.[21,22]

The EQ-5D (3L version) is a self-administered questionnaire consisting of one question in each of five dimensions pertaining to specific health domains (mobility, self-care, pain, usual activities, and anxiety/depression). Patients rate each dimension with three possible responses: no problems, some problems, or extreme problems. EQ-5D health domain scores are converted into a single summary utility index, or summary health index, using one of the available EQ-5D value sets. For this study, UK population sample weights were used. Summary health index scores range from −0.594 to 1.0 with higher scores representing better health states. Based on data from eight longitudinal studies with conditions including leg ulcers, back pain, rheumatoid arthritis, irritable bowel syndrome, limb reconstruction, acute myocardial infarction, osteoarthritis, and chronic obstructive pulmonary disease, a predefined change in summary health index score of less than −0.074 points at a given time point from baseline defined the threshold for worsening.[23] The EQ-5D also has a visual analogue scale (VAS) rating of current health state ranging from 0 (worst imaginable health state) to 100 (best imaginable health state). Change in EQ-5D VAS score less than −5.5 at a given time point from baseline represents the threshold for worsening when examined among a population of patients with metastatic colon cancer.[24]

Anchor-based evaluation

SELECT data provide the opportunity to evaluate the RD of the MSIS-29 PHYS using several anchors, including EDSS change, SF-12 PCS, EQ-5D, and EQ-5D VAS. Because these anchor instruments are most closely related to the MSIS-29 PHYS, and because most MS disease-modifying therapies slow, but do not reverse, disability progression, only the MSIS-29 PHYS RD for worsening was evaluated using SELECT data. The term 'responder' requires some clarification in the context of this analysis. For the study presented here, we use 'responder' in the psychometric sense of having a change on the outcome of interest, progression of disability; as opposed to 'responder' in the clinical sense of experiencing some improvement in response to a treatment. As data from all subjects in the trial who had confirmed disability progression were treated equally in the estimation of the RD, the treatment effect of DAC HYP did not impact this analysis.

Longitudinal correlation was used to confirm that each of the proposed anchors had good correlation with the MSIS-29 PHYS. Longitudinal correlations of MSIS-29 PHYS change scores with change in EDSS and change scores from the EQ-5D summary health index, EQ-5D VAS, and SF-12 PCS score were examined at each post-baseline time point. Change scores at 12, 24, and 52 weeks identified which measures change together with the MSIS-29, and a change score correlation >0.30 was preferred.[25]

Patients were dichotomized as responders or non-responders based on the predefined RD of an anchor measure. Responders in this study referred to patients who deteriorated by at least a predefined threshold value of an anchor measure (i.e. their score worsened on the anchor measure; these patients were classified as non-responders in relation to daclizumab treatment). The primary anchor of interest was the EDSS. Sustained disability progression in SELECT could only be confirmed at a scheduled visit during which EDSS assessment was made (i.e. 12, 20, 24, 36, 48, and 52 weeks). However, measurement of MSIS-29 only occurred at pre-specified time points (12, 24, and 52 weeks) that did not always occur simultaneously with confirmation of progression using the EDSS. Therefore, assessment of changes in MSIS-29 scores from baseline used four scenarios: (1) following onset of progression; (2) at or after confirmation of progression; (3) at or before confirmation of progression; and (4) within ±4 weeks of the confirmed progression.

Table 1. Baseline demographics and characteristics from SELECT.

Characteristic*	Placebo (n = 196)	DAC HYP 150 mg (n = 201)	DAC HYP 300 mg (n = 203)
Age, y	36.9 (9.0)	35.2 (9.1)	35.4 (8.6)
Female, *n* (%)	123 (62.8)	136 (67.7)	132 (65.0)
Disease duration, y	4.1 (5.3)	4.5 (5.0)	3.8 (4.0)
Number of relapses in past year	1.3 (0.6)	1.4 (0.7)	1.3 (0.7)
EDSS score	2.7 (1.2)	2.8 (1.1)	2.6 (1.2)
Number of Gd+ lesions, mean (median)	2.0 (0)	2.1 (1.0)	1.4 (0)
Number of T2 hyper-intense lesions, mean (median)	40.0 (30.0)	44.8 (36.0)	35.8 (28.5)
MSIS-29			
Physical impact subscale	26.3 (22.0)	24.7 (20.2)	24.0 (19.5)
Psychological impact subscale	29.5 (22.5)	28.6 (21.5)	29.6 (20.7)
SF-12			
PCS	42.5 (10.0)	42.9 (9.9)	43.1 (9.0)
MCS	46.4 (10.2)	46.1 (11.5)	45.5 (11.0)
EQ-5D			
VAS	71.2 (18.3)	72.0 (17.4)	72.1 (18.1)
Summary Health Index	0.7 (0.2)	0.7 (0.2)	0.7 (0.2)

DAC HYP: daclizumab high-yield process; EDSS: Expanded Disability Status Scale; EQ-5D: EuroQol 5-Dimensions; Gd+: gadolinium-enhancing; MCS: mental component summary; MSIS-29: Multiple Sclerosis Impact Scale; PCS: physical component summary; SF-12: Short-Form Health Survey-12; VAS, visual analogue scale. *Values are reported as mean (standard deviation), except where noted.

In this analysis, RD was calculated using mean or median change scores in MSIS-29 PHYS scores at 12, 24, and 52 weeks (for the EQ-5D summary health index, EQ-5D VAS, and SF-12 PCS anchors) among patients who met the predefined anchor-specific thresholds for responders (i.e. those with worsening health-related quality of life).

Distribution-based evaluation

Two strategies were used to calculate RD using the distribution-based approach. The first calculated RD as the ES for the MSIS-29 PHYS score as the mean change score/SD at baseline.[21,26] The RD was the value that corresponded to an ES of 0.3 or 0.5. The second strategy estimated the RD as 1 SEM (SEM = baseline SD × [√1 − reliability]).[27] The SEM is expressed in the original metric of the instrument, and change beyond 1 SEM has demonstrated correspondence with an important change in several other chronic diseases.[26–29] The reliability of the MSIS-29 PHYS was assessed by the intra-class correlation coefficient from an intercept-only random effects model, which was used in the SEM calculation.

The responder analysis comparing the proportions of patients with change in MSIS-29 PHYS score considered RD as a treatment arm at all time points. The Cochran–Mantel–Haenszel test was used to identify differences in proportion between groups.

Analyses were completed using SAS v9.2 (SAS Institute Inc., Cary, NC, USA).

Results

Baseline characteristics for the ITT efficacy population from SELECT (*n* = 600) are shown in Table 1.[17] Most (65.2%) patients were female and mean EDSS score (2.6–2.8) was similar among treatment groups. The current analysis included only patients with ≥1 baseline and 1 post-baseline observed PRO score (placebo, *n* = 189; DAC HYP 150 mg, *n* = 193; DAC HYP 300 mg, *n* = 191). When the MSIS-29 PHYS had <10 items missing, the score was imputed using the mean of the completed items.[17] Relatively few (<5%) patients were missing PRO data.

For all selected anchors, magnitudes of the cross-sectional correlations at baseline for the MSIS-29 PHYS were >0.50. All correlations (cross-sectional and longitudinal) had the expected directionality. Longitudinal correlations were consistent over time points (Table 2). Statistically significant moderate correlations (>0.30) between the MSIS-29 PHYS and EQ-5D or SF-12 PCS were observed, with the best correlations being between the MSIS-29 PHYS and the EQ-5D summary health index or SF-12 PCS. Correlations with EDSS change scores were weaker ($0.14 \geq r \leq 0.24$; $p < 0.05$).

Table 2. Longitudinal correlations between the MSIS-29 PHYS, EDSS, and SELECT HRQOL measures.

Change from baseline	EDSS	EQ-5D summary health index	EQ-5D VAS	SF-12 PCS
MSIS-29 PHYS				
To week 12	0.24[a]	−0.37[a]	−0.31[a]	−0.37[a]
To week 24	0.14[b]	−0.41[a]	−0.35[a]	−0.46[a]
To week 52	0.22[a]	−0.35[a]	−0.37[a]	−0.44[a]

[a]$p < 0.0001$.
[b]$p < 0.05$.
EDSS: Expanded Disability Status Scale; EQ-5D: EuroQol 5-Dimensions; HRQOL: health-related quality of life; MCS: mental component summary; MSIS-29: Multiple Sclerosis Impact Scale; PCS: physical component summary; PHYS: physical impact subscale; SF-12: Short-Form Health Survey-12; VAS: visual analogue scale.

Table 3. Anchor-based approach results for determining the RD of the MSIS-29 PHYS.

Measure	*n*	RD	Mean	Median
EDSS-1[a]	84	Progression[e]	8.21	7.50
EDSS-2[b]	38	Progression[e]	5.98	7.50
EDSS-3[c]	48	Progression[e]	7.25	6.88
EDSS-4[d]	25	Progression[e]	6.46	7.50
EQ-5D summary health index	386	0.074	7.2	6.25
EQ-5D VAS	379	5.5	6.05	3.75
SF-12 PCS	410	3	7.07	5.00
SF-12 PCS	293	5	8.55	6.25
SF-12 PCS	240	6	9.49	7.50

[a]EDSS-1: All MSIS-29 PHYS scores following the start of progression were used.
[b]EDSS-2: Only the MSIS-29 PHYS score at or after the confirmation of progression was used.
[c]EDSS-3: Only the MSIS-29 PHYS score at or before the confirmation of progression was used.
[d]EDSS-4: Only the MSIS-29 PHYS score within ± 4 weeks of the confirmed progression was used.
[e]Progression was confirmed.
EDSS: Expanded Disability Status Scale; EQ-5D: EuroQol 5-Dimensions; MSIS-29: Multiple Sclerosis Impact Scale; PCS: physical component summary; PHYS: Physical Impact Subscale; RD: responder definition; SF-12: Short-Form Health Survey-12; VAS: visual analogue scale.

Using the anchor-based approach, mean and median RD values were calculated for each anchor (Table 3). Anchor-based approaches yielded mean, median, and mode RD values of 6.91, 7.14, and 7.50, respectively (range 3.75–9.48). This mean is the average of all RDs based on mean or median changes (Table 3). Similar calculations were performed for median and mode. Three of four EDSS anchor scenarios resulted in a median RD of 7.50; the remaining EDSS anchor scenario had a median RD of 6.88 (Table 3). Sensitivity analyses of patients with EDSS change equal to 1.00 or 1.50 produced similar change results (data not shown). Distribution-based RD estimates were 8.05 for SEM, 6.24 for an ES of 0.3, and 10.40 for an ES of 0.5.

Based on both anchor- and distribution-based approaches, RDs of the MSIS-29 PHYS range from 3.75 to 10.34. An RD of 7.50 was selected as the most appropriate RD threshold for physical worsening for the following reasons: (1) a median of 7.50 was demonstrated in three of four EDSS scenarios (primary anchor of interest); (2) anchor-based methods that used a 6.0-point worsening in SF-12 PCS score provided additional support for an RD of 7.50 as the most appropriate threshold for the MSIS-29; (3) results from distribution-based methods (0.5 ES and SEM) suggested this was an appropriate RD; and (4) other change levels (e.g. 7.00 or 7.25) cannot represent the RD threshold for an individual on the MSIS-29 because 1.25 is the smallest increment of an individual patient's scores that represents a single unit change of a single item on the MSIS-29 PHYS. Score changes other than 1.25 can only be achieved in patients who provide incomplete data on the MSIS-29; this was not only quite rare, but violates the principles of good clinical practice in seeking complete PRO responses at each visit and should not be the basis of an RD.
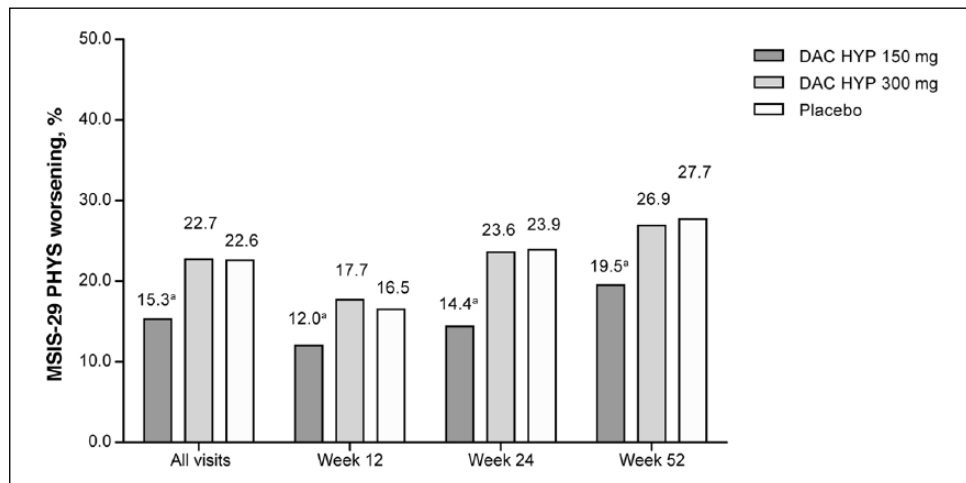
Figure 1. Responder analysis using data from SELECT and applying a responder definition of 7.50.
[a]$p < 0.01$ compared with placebo or DAC HYP 300 mg.
DAC HYP: daclizumab high-yield process; MSIS-29: Multiple Sclerosis Impact Scale; PHYS: physical impact subscale.

Therefore, an RD of 7.50 represents the best choice from the achievable thresholds (e.g. 6.25, 7.50, 8.75, etc.) for MSIS-29 PHYS change scores.

Using an RD value of 7.50, the proportion of patients in SELECT with MSIS-29 PHYS worsening was significantly lower at each follow-up visit (weeks 12, 24, and 52) in the DAC HYP 150 mg–treated group compared with DAC HYP 300 mg– or placebo-treated groups (Figure 1). At week 52, 19.5% of DAC HYP 150 mg–treated patients had MSIS-29 PHYS worsening compared with 26.9% of DAC HYP 300 mg–treated patients or 27.7% of placebo-treated patients ($p < 0.01$ for both vs. placebo; Figure 1).

Discussion
Our finding of an RD of 7.50 for the MSIS-29 PHYS in patients with RRMS and physical worsening in the clinical trial setting adds to the interpretability of future MSIS-29 PHYS results. The RD of 7.50 from our analysis is comparable with the RD of 7.00 reported for the MSIS-29 PHYS in a community-based population with EDSS scores of 0–5.0, and 8.00 for patients with EDSS scores of 5.5–8.0.[16] As in our study, a weaker correlation between the MSIS-29 PHYS in the EDSS range of 0.0–5.0 was noted with report of a possible floor effect resulting from symptoms that are addressed in EDSS score from 0 to 4.0. In this range, assessment of impairment is more focused on neurological rather than functional impairment. This effect may contribute to the statistically significant but weak correlation observed between the MSIS-29 PHYS and EDSS. The difference between an RD of 7.00 and our finding of 7.50 could be due to

the selection of patients included in each analysis (i.e. inclusion of patients with progressive forms of MS in the community study). Moreover, an RD of 7.00 is not achievable given that the individual state change of the MSIS-29 is 1.25 units. While the similarity between the RD of 7.50 in the current analysis and that of Costelloe et al.[16] is similar and thus reassuring, it cannot be assumed that it can be applied to all RRMS populations. Confirmatory work by other investigators using their own clinical trial population is encouraged.

When the RD of 7.50 was applied to the SELECT patient population, the proportion of patients with physical worsening (MSIS-29 PHYS) was significantly lower at all time points in those treated with DAC HYP 150 mg compared with those treated with placebo, which is consistent with other PROs from the trial. In SELECT, significant improvements from baseline in MSIS-29 PHYS, SF-12 PCS and EQ-5D VAS scores were observed for the DAC HYP 150-mg group at 24 and 52 weeks compared with placebo; only the EQ-5D VAS at 52 weeks showed significant improvement for the DAC HYP 300-mg group.[30] The lack of a dose response for DAC HYP 300 mg in the current analysis is consistent with analyses evaluating mean change from baseline in PRO outcomes in SELECT.[30]

Strengths of the current analysis include use of multiple anchors and low rates of discontinuation in SELECT. Although an anchor that directly measured patient-perceived change in physical functioning would have been preferable, this was not an assessment in SELECT. Some patients with an MSIS-29

PHYS score below the RD of 7.50 may still have clinically significant worsening. Another limitation of this analysis is that SELECT included only patients with RRMS with baseline EDSS scores of 0–5.0; therefore, these findings may not be generalizable to patients with more advanced or progressive disease. Also, the EQ-5D RD anchors (summary health index score less than −0.074 points and VAS score less than −5.5) were not derived from studies that included patients with RRMS. This analysis was based on an EDSS change confirmed over 3 months. Larger trials with longer study duration may be able to estimate the RD based on EDSS progression confirmed at 6 months. A final limitation is that measurement time points were based on a predetermined protocol driven schedule, and therefore were not likely to be administered at the time of an attack. Future research into meaningful methods to obtain relevant PRO assessment at the time of a relapse or identified disease progression would provide additional insights to understanding important changes over time.

The MSIS-29 is a reliable validated responsive measure of the impact of MS from the patient's perspective.[4,6,8–11] The findings of this analysis indicate that worsening on the MSIS-29 PHYS ≥7.50 is a reasonable and practical threshold for identifying patients with RRMS who have experienced a clinically significant change in the physical impact of MS. PROs can provide unique information on the effects of treatment, thus, establishing an RD for the MSIS-29 PHYS provides the basis for its use in future clinical trials of RRMS and clinical practice. Prospective studies will need to determine whether they can be replicated and extended to other MS populations.

Conflict of interest

Glenn Phillips, Rossella Medori, and Jacob Elkins are full-time employees of Biogen Idec.
Shien Guo, Kathleen W Wyrwich, and Arman Altincatal are full-time employees of Evidera.
Linda Wagner, who is a full-time employee of Excel Scientific Solutions, part of the Envision Pharma Group, provided medical writing and editorial support in the development of this manuscript and this work was funded by Biogen Idec.

References

1. US Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. *Fed Regist* 2009; 74: 65132–65133.

2. US Food and Drug Administration. Draft guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims; availability. *Fed Regist* 2006; 71: 5862–5863.

3. Wyrwich KW, Norquist JM, Lenderking WR, et al. Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res* 2013; 22: 475–483.

4. Hobart J, Lamping D, Fitzpatrick R, et al. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. *Brain* 2001; 124: 962–973.

5. Gray O, McDonnell G and Hawkins S. Tried and tested: the psychometric properties of the multiple sclerosis impact scale (MSIS-29) in a population-based study. *Mult Scler* 2009; 15: 75–80.

6. Hobart JC, Riazi A, Lamping DL, et al. Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome. *Health Technol Assess* 2004; 8(iii): 1–48.

7. Hobart JC, Riazi A, Lamping DL, et al. How responsive is the Multiple Sclerosis Impact Scale (MSIS-29)? A comparison with some other self report scales. *J Neurol Neurosurg Psychiatry* 2005; 76: 1539–1543.

8. Hoogervorst EL, Zwemmer JN, Jelles B, et al. Multiple Sclerosis Impact Scale (MSIS-29): relation to established measures of impairment and disability. *Mult Scler* 2004; 10: 569–574.

9. McGuigan C and Hutchinson M. The multiple sclerosis impact scale (MSIS-29) is a reliable and sensitive measure. *J Neurol Neurosurg Psychiatry* 2004; 75: 266–269.

10. Ramp M, Khan F, Misajon RA, et al. Rasch analysis of the Multiple Sclerosis Impact Scale MSIS-29. *Health Qual Life Outcomes* 2009; 7: 58. Published online June 22, 2009. doi: 10.1186/1477-7525-7-58.

11. Riazi A, Hobart JC, Lamping DL, et al. Multiple Sclerosis Impact Scale (MSIS-29): reliability

and validity in hospital based samples. *J Neurol Neurosurg Psychiatry* 2002; 73: 701–704.

12. Costelloe L, O'Rourke K, McGuigan C, et al. The longitudinal relationship between the patient-reported Multiple Sclerosis Impact Scale and the clinician-assessed Multiple Sclerosis Functional Composite. *Mult Scler* 2008; 14: 255–258.

13. Ravnborg M, Blinkenberg M, Sellebjerg F, et al. Responsiveness of the Multiple Sclerosis Impairment Scale in comparison with the Expanded Disability Status Scale. *Mult Scler* 2005; 11: 81–84.

14. Riazi A, Hobart JC, Lamping DL, et al. Evidence-based measurement in multiple sclerosis: the psychometric properties of the physical and psychological dimensions of three quality of life rating scales. *Mult Scler* 2003; 9: 411–419.

15. Giordano A, Pucci E, Naldi P, et al. Responsiveness of patient reported outcome measures in multiple sclerosis relapses: the REMS study. *J Neurol Neurosurg Psychiatry* 2009; 80: 1023–1028.

16. Costelloe L, O'Rourke K, Kearney H, et al. The patient knows best: significant change in the physical component of the Multiple Sclerosis Impact Scale (MSIS-29 physical). *J Neurol Neurosurg Psychiatry* 2007; 78: 841–844.

17. Gold R, Giovannoni G, Selmaj K, et al. Daclizumab high-yield process in relapsing-remitting multiple sclerosis (SELECT): a randomised, double-blind, placebo-controlled trial. *Lancet* 2013; 381: 2167–2175.

18. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; 33: 1444–1452.

19. Goldman MD, Motl RW and Rudick RA. Possible clinical outcome measures for clinical trials in patients with multiple sclerosis. *Ther Adv Neurol Disorders* 2010; 3: 229–239.

20. Ware J, Jr, Kosinski M and Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996; 34: 220–233.

21. Norman GR, Sloan JA and 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.

22. Robinson D Jr, Zhao N, Gathany T, et al. Health perceptions and clinical characteristics of relapsing-remitting multiple sclerosis patients: baseline data from an international clinical trial. *Curr Med Res Opin* 2009; 25: 1121–1130.

23. Walters SJ and Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005; 14: 1523–1532.

24. Mathias SD, Pritchard ML, Colwell HH, et al. What is the minimal clinically important difference and responsiveness of a patient-reported outcome questionnaire for metastatic colorectal cancer? *Ann Oncol* 2006; 17: Abstract no. 349P.

25. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; 61: 102–109.

26. Cella D, Eton DT, Fairclough DL, et al. What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. *J Clin Epidemiol* 2002; 55: 285–295.

27. Wyrwich KW, Nienaber NA, Tierney WM, et al. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999; 37: 469–478.

28. Wyrwich KW, Tierney WM and Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying changes in health-related quality of life. *J Clin Epidemiol* 1999; 52: 861–873.

29. Wyrwich KW, Tierney WM and Wolinsky FD. Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire. *Qual Life Res* 2002; 11: 1–7.

30. Selmaj K, Havrdova E, Gold R, et al. Daclizumab HYP monotherapy improved health-related quality of life parameters in relapsing-remitting multiple sclerosis: findings of the SELECT trial. *Neurology* 2012; 78: P07.098.Pistotatet qunt et fugit volluptios