ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Preference-Based Assessments

Comparative Responsiveness of Preference-Based Health-Related Quality of Life, Social Care, and Well-Being Measures in the Context of Multiple Sclerosis

Elizabeth Goodwin, PhD, Amy Heather, MSc, Nia Morrish, MSc, Jenny Freeman, PhD, Kate Boddy, MSc, Sarah Thomas, PhD, Jeremy Chataway, PhD, Rod Middleton, PhD, Annie Hawton, PhD

ABSTRACT

Objectives: To provide evidence on the responsiveness of social care and well-being preference-based measures (PBMs) compared with health-related quality of life PBMs in the context of multiple sclerosis (MS).

Methods: The ICEpop CAPability measure for Adults (ICECAP-A) and Adult Social Care Outcomes Toolkit (ASCOT) were completed online in September 2019, March 2020, September 2020, via the UK MS Register. Responses were linked to EQ-5D-3L and MS Impact Scale-8 Dimensions (MSIS-8D) values, and to MS Walking Scale-12, Hospital Anxiety and Depression Scale (HADS), and Fatigue Severity Scale scores. Responsiveness was assessed in relation to minimal important differences on MS Walking Scale-12, Hospital Anxiety and Depression Scale, and Fatigue Severity Scale between time points, using mean change scores, *t* tests, standardized effect sizes, standardized response means, and multivariable regression analyses.

Results: Data from 1742 people with MS were available for analysis. When using standardized values, MSIS-8D showed the greatest responsiveness and EQ-5D-3L the least. In contrast, when absolute utility values were used, EQ-5D-3L performed similarly to MSIS-8D and better than ICECAP-A and ASCOT. Standardized regression analyses indicated the MSIS-8Ds to be the most responsive, followed by the ASCOT, ICECAP-A, and EQ-5D-3L.

Conclusions: The ICECAP-A, ASCOT, and MSIS-8D were more responsive than the EQ-5D-3L in the context of MS when compared using standardized scores. The increased responsiveness of EQ-5D-3L when absolute values were used seems an artefact of the wide-ranging scale of this measure. This illustrates how the maximum potential range of values for a given PBM tariff could influence whether an intervention is found to be cost-effective.

Keywords: EQ-5D, health-related quality-of-life measures, preference-based measures, QALY measures, responsiveness, social-care-related quality-of-life measures, well-being measures.

VALUE HEALTH. 2026; 29(3):476–487

Highlights

- Evidence is limited for the responsiveness of preference-based measures (PBMs) of health-related quality of life, well-being, or social-care-related quality of life in multiple sclerosis.
- This article compares the responsiveness of 5 PBMs: EQ-5D-3L, MSIS-8D, MSIS-8D-P, ICECAP-A, and ASCOT. When values were standardized, MSIS-8D and MSIS-8D-P showed the greatest responsiveness, whereas EQ-5D-3L showed the least. Using absolute values, EQ-5D-3L and the MSIS-8D measures showed the greatest responsiveness.
- Comparative responsiveness can vary depending upon the metric(s) used to judge responsiveness. These findings illustrate how the maximum range of values for a PBM could influence whether an intervention is found to be cost-effective and suggest that careful interpretation is required to ensure nonhealth impacts are not eclipsed by "larger" health impacts.

Introduction

The funding of health and social care treatments in the UK largely depends on demonstrating their cost-effectiveness,[1] which is typically assessed by weighing up the costs of treatments against the benefits they provide.[2,3] Concerns have been raised, however, about the ability of the outcome measures typically used in cost-effectiveness analysis to capture all the benefits of interventions that are experienced by people with multiple sclerosis (MS).[4-9] This highlights the importance of responsiveness, ie, the ability of a measurement tool to accurately detect and quantify a change over time in the construct of interest when it has occurred.[10] If scores on an outcome measure do not sufficiently alter as a result of significant events, important treatment effects may be missed. Such issues have hampered funding decisions[11-13] and may have limited the services that are available to people with MS. Consequently, in 2012, the National Institute for Health and Care Excellence (NICE) called for improvements to the framework for assessing the cost-effectiveness of treatments for MS.[13]

Much of the research in this area has focused on measures of health-related quality of life (HRQL).[14] In the United Kingdom, NICE[15] recommends the EQ-5D, a generic preference-based measure (PBM) of HRQL with 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.[16] Concerns regarding the responsiveness of generic PBMs to MS treatment effects[4-9] led to the development of a condition-specific PBM of HRQL: the Multiple Sclerosis Impact Scale 8 Dimensions (MSIS-8D), which is derived from the Multiple Sclerosis Impact Scale-29 (MSIS-29)[17] and includes HRQL

dimensions of particular importance to people with MS: physical, social, mobility, daily activities, fatigue, emotion, cognition, and depression.[18] Both measures have a tariff of utility values based on the preferences of a UK general population sample, enabling the calculation of quality-adjusted life-years (QALYs).[16,19] In addition, the MSIS-8D-P provides a tariff based on the preferences of a UK sample of people with MS, for the same descriptive system.[20]

HRQL measures are designed to assess interventions that aim primarily to improve health. However, there is increasing interest in broadening the evaluative space to incorporate "benefits beyond health."[21] A key development in the United Kingdom was the extension of the remit of NICE in 2013 to include social care. NICE recommends 2 instruments to capture the nonhealth effects of social care interventions.[15] These generic measures each have a tariff of values, based on the preferences of a UK general population sample, enabling the calculation of equivalents to the health-related QALY.[22,23]

For interventions with social care outcomes, NICE recommends the Adult Social Care Outcomes Toolkit (ASCOT).[22] The ASCOT measures social care-related quality of life (SCRQL) and includes the following dimensions: control over daily life, personal cleanliness and comfort, food and drink, personal safety, social participation and involvement, occupation, accommodation cleanliness and comfort, and dignity.

For interventions with well-being outcomes, NICE recommends the ICEpop CAPability measure for Adults (ICECAP-A).[23] The ICECAP-A measures "capability well-being," ie, being able to do and be the things in life that matter to individuals.[24] Its attributes cover the following: attachment, security, role, enjoyment, and control.

The symptoms of MS, and the variety of interventions and services used by people with MS, have meaningful impacts on people with MS beyond their health and HRQL, including role functioning, employment, self-efficacy, and independence.[25-29] PBMs of HRQL do not directly capture beneficial effects of interventions and services on these factors.[6,29] This suggests that instruments designed to measure broader aspects of quality of life and well-being may be more responsive to these nonhealth effects. This research aimed to explore whether social care and well-being PBMs (SCW-PBMs) are more responsive than HRQL-PBMs in the context of MS by assessing and comparing the responsiveness of the ICECAP-A, ASCOT, EQ-5D-3L, and MSIS-8D, using data from the UK MS Register (UKMSR).

Methods

Patient-Reported Outcome Measures and Minimal Important Differences

The responsiveness of the PBMs was assessed in relation to minimal important differences (MIDs) on 3 patient-reported outcome measures (PROMs): the Multiple Sclerosis Walking Scale-12 (MSWS-12),[30] the Hospital Anxiety and Depression Scale (HADS)[31] and the Fatigue Severity Scale (FSS).[32] These measures form part of the UKMSR's routine data collection and represent frequently experienced MS symptoms that affect HRQL[27,28] and broader aspects of well-being.[29] Higher scores on these measures indicate greater symptom severity; therefore, a positive change indicates a deterioration and a negative change indicates an improvement. Published MIDs for each PROM were sourced from the literature. When there was more than 1 possible MID, the MID for use in the analysis was agreed through discussion among the research team.

The MSWS-12 version 2.0 is a validated measure of walking ability, consisting of 3 items with 3 response levels and 9 items with 5 response levels. Raw scores are recalibrated to a 0 to 100 point scale. The MID of ≥8 points was used for analyses.[33-35]

The HADS is a validated measure with separate subscales for depression and anxiety.[31] Each subscale comprises 7 items with 4 response levels, giving a total subscale score range of 0 to 21. Estimates of the MID range from 1.5 to 2.5 points.[36-38] A MID of 2.0 was used for analysis of each HADS subscale.

The FSS is a 9-item measure of the severity and impact of fatigue. Total scores are expressed as the average score across the items, giving a range from 1 to 7 points. Estimates for FSS MIDs range from 0.45 to 1.1 points.[39,40] A MID of 1.0 was used for this analysis.

Survey of People With MS

The UKMSR administers an online survey every 6 months, which is completed by approximately 5000 people with MS. The survey includes PROMs alongside sociodemographic and clinical information. It can also host study-specific questionnaires, which can be anonymously linked to its existing data.[41]

The ICECAP-A and the ASCOT were administered 3 times over an 18-month period (September 2019, March 2020, and September 2020). This was alongside the UKMSR's regular surveys at the same time points, which collected data on: EQ-5D-3L, MSIS-29, MSWS-12, HADS, FSS, age, gender, type of MS, and date of diagnosis. Values for the PBMs were estimated using published UK tariffs: EQ-5D-3L,[16] MSIS-8D based on general population preferences,[19] MSIS-8D-P based on preferences of people with MS,[20] ICECAP-A,[23] and ASCOT.[22]

Following their standard procedures, the UKMSR emailed all members of the register, inviting them to take part and providing them with a link to the additional questionnaires. Previous surveys hosted by the UKMSR suggested that we could expect a sample of approximately 2000 at each time point.

No personal data were collected that could identify individuals. All data were accessed via an encrypted secure desktop connection. Data were analyzed using Stata17, with do-files available on GitHub (https://github.com/exeterhems/hums_responsiveness) and Zenodo.[42] Ethical approval for the UKMSR has been provided by the South West Central Bristol Research Ethics Council 21/SW/0085.

Responsiveness Analysis

The responsiveness of the PBMs was assessed in relation to MIDs on each of the PROMs (MSWS-12, HADS depression, HADS anxiety, and FSS). A series of complete case analyses was undertaken using data from all respondents who provided at least 2 consecutive responses to at least 1 of the PBMs over the 3 time points.

Based on MIDs in the total or subscale score of each PROM compared with the previous time point, the following groups were identified:

1. individuals with a MID indicative of an improvement
2. individuals with a MID indicative of a worsening
3. individuals with no MID

Hypotheses regarding these groups, ceteris paribus, were that there would be

Table 1. Illness-related events.

Change in treatments and services since previous time point	Response category
Receiving any treatments for MS	Yes/No
Receiving a new drug treatment for MS symptoms	Yes/No/NA
Stopping a drug treatment for MS symptoms Did you stop the medicine/drug treatment because …	Yes/No/NA It made things worse? It made no difference? Symptoms improved? Other reason
Receiving a new or existing disease-modifying treatment Did you …	Yes/No/NA Take a DMT throughout? Start a DMT? Stop a DMT? Switch DMTs?
Receiving a new non-drug treatment or therapy for MS	Yes/No/NA
Stopping a non-drug treatment or therapy for MS Did you stop the treatment/therapy because …	Yes/No/NA It made things worse? It made no difference? Symptoms improved? Other reason
Change in NHS or social care services Overall are the services you receive now better, the same or worse?	Yes/No/NA Better Same Worse
Changed in self-funded services or support Overall are the services you receive now better, the same or worse?	Yes/No/NA Better Same Worse

Changes in employment since previous time point	Response category
Been in paid work at any point in the last 6 months	Yes/No
Permanently left employment due to MS Did you want to leave your job?	Yes/No/NA Yes No
Changed job due to MS? Did you want to change your job?	Yes/No/NA Yes No
Reduced working hours due to MS Did you want to reduce your hours?	Yes/No/NA Yes No
Returned to work after having been off for more than 4 weeks due to MS Did you go back to work because …	Yes/No/NA You needed to You wanted to

Experiences of MS since previous time point	Response category
Change of diagnosis from RRMS to SPMS	Yes/No/NA
Experiencing new MS symptoms	Yes/No
Experiencing new complications or side-effects due to MS treatment	Yes/No/Don't know/NA
Experiencing an MS relapse	Yes/No/NA
Additional questions on relapses	
How many relapses have you had in the last 6 months?	One Two Three or more
How long did the first relapse last? By this we mean how long was it until the symptoms settled down or disappeared?	48 hours 1 week 1 month Over 1 month Ongoing
Were you admitted to hospital as a result of this relapse?	Yes/No

continued on next page

Table 1. Continued

Experiences of MS since previous time point	Response category
Did you spend any time off work as a result of having this relapse?	Yes/No/NA
Did having this relapse limit any other everyday activities?	Yes/No
Did you have a second relapse in the last 6 months?	Yes/No

DMT indicates disease modifying treatment; MS, multiple sclerosis; NA, not applicable; NHS, National Health Service.

1. a significant increase in PBM values
2. a significant decrease in PBM values
3. no significant change in PBM values

This gave a total of 16 groups in which a change in PBM values was hypothesized: 1 improvement group and 1 deterioration group for each of the 4 PROMs (8 groups), for time point 1 to 2 and for time point 2 to 3 (16 groups).

Paired *t* tests were undertaken to assess the statistical significance ($P < .05$) of differences between PBM values from time point 1 to time point 2, and from time point 2 to time point 3, for each respondent group. Results were adjusted for multiple testing using Bonferroni correction. The magnitude of changes in PBM values between time points was assessed using the absolute size of the change, as well as standardized effect sizes (SES) and standardized response means (SRM) to enable comparison across measures on a standardized scale. SES were calculated as the mean change divided by the standard deviation of the earlier mean score (Cohen's *d* statistic), and interpreted as no (<0.2), small (0.2), moderate (0.5), or large (0.8) effect.[43] SRMs were calculated by dividing the mean change by the standard deviation of the mean change score,[44] and were directly compared between PBMs.[45] The 5 PBMs were ranked for each respondent group according to the absolute size of the SRM. The mean and modal average ranks across all respondent groups were then calculated for each PBM. The PBMs were also ranked for each respondent group according to the absolute size of the nonstandardized mean change score, and their mean and modal average ranks across all groups were calculated.

The comparison of changes in PBM values was undertaken using the following metrics:

- Number of respondent groups in which there was a statistically significant difference in PBM values between time points.
- Number of respondent groups in which there was at least a small effect size.
- Average rank in terms of the absolute size of the SRM.
- Average rank in terms of the absolute size of the mean change score.

Regression Responsiveness Analyses

Multivariable regression analyses were used to compare the responsiveness of the PBMs by exploring independent relationships between change on each PROM and change on each PBM. A separate regression model was run for each PBM.

Two sets of regression analyses were undertaken using (1) change scores on the continuous scale of the PROMs and (2) changes on the PROMs coded as "no MID" (reference case), "MID indicative of an improvement," and "MID indicative of a worsening." Continuous scores on the PROMs and the PBMs were standardized to a mean of 0 and a standard deviation of 1 to enable direct comparison.

Because panel data were used, the appropriateness of fixed-effect and random-effects models was investigated. The Hausman test indicated that random-effects should be used over fixed-effects. However, the Breusch-Pagan Lagrange multiplier test found no evidence of significant differences across time points, indicating that random effects was inappropriate. Therefore, a multivariable pooled OLS regression was run with standard errors clustered by respondent. The data failed the assumption of normality of residuals, requiring a choice between conducting a robust regression or a regression with clustered standard errors. The latter was chosen, supported by the observation that the *P* values were larger and less significant when clustering, indicating this was the more conservative choice.

Illness-Related Events

The relevance of the PROMs for assessing the responsiveness of SCW-PBMs is questionable. The PROMs were designed to measure individual symptoms, which are more closely related to health status than SCRQL or well-being. The types of interventions for which the SCW-PBMs are recommended are unlikely to directly affect specific symptoms. Therefore, an Illness-Related Events (IRE) questionnaire was designed for completion by people with MS. Significant IREs, which people with MS describe as affecting their quality of life and well-being, were identified via a systematic "review of reviews" of the relevant literature. This resulted in a "longlist" of IREs. The Health Economics and Multiple Sclerosis (HEMS) Patient Involvement Group[46] worked with the researchers to reduce this to a shortlist of the most impactful events, amenable for assessing PBM responsiveness and for inclusion in the UKMSR surveys. The resulting IRE questionnaire was discussed in detail with the coapplicant team, explored in cognitive interviews with people with MS, and piloted online with 50 members of the UKMSR. The questionnaire was administered alongside the UKMSR surveys in September 2019, March 2020, and September 2020 and asked people with MS to report whether or not they had experienced each of the specific IREs over the previous 6 months (Table 1). The responsiveness of the PBMs in relation to these IREs was analyzed using the same methods as with the PROMs, with groups of respondents determined by whether they had experienced a particular IRE.

Public and Patient Involvement (PPI)

The HEMS Patient Involvement Group[46] were integral to this research, meeting with the project team 6 times over the life of the study, with multiple additional contacts via email. The Group's involvement included, but was not limited to, shaping the focus of the research and the application, helping the project team to understand the relevance of the research from the perspective of people with MS, giving their views on the study materials, interpreting the findings, considering the implications for the MS community, and guiding next steps in terms of policy, practice, and research.

Table 2. Descriptive statistics for all respondents who provided at least 2 consecutive responses to at least 1 PBM (*N* = 1742) and Spearman correlations between measures at time point 1 to assess suitability for anchor-based responsiveness analysis.

Respondent characteristic	*N*	%	Mean	SD	Observed range	Theoretical range	UK general population
Time point 1							
Gender - Female	1288	74.07	-	-	-	-	51.67%
Gender - Male	451	25.93	-	-	-	-	48.33%
MS type - RRMS	745	49.24	-	-	-	-	NA
MS type - SPMS	296	19.56	-	-	-	-	NA
MS type - PPMS	328	21.68	-	-	-	-	NA
MS type - benign	26	1.72	-	-	-	-	NA
MS type - Unknown	118	7.80	-	-	-	-	NA
Age	1740	-	54.630	11.053	21 to 85	-	49.143
MSWS-12	1717	-	61.905	37.763	0 to 100	0 to 100	NA
HADS depression score	1717	-	6.893	4.569	0 to 21	0 to 21	3.68
HADS anxiety score	1717	-	6.710	4.621	0 to 21	0 to 21	6.14
FSS score	1713	-	4.956	1.505	1 to 7	1 to 7	NA
EQ-5D value	1725	-	0.567	0.310	−0.594 to 1	−0.594 to 1	0.856
MSIS-8D value	1722	-	0.617	0.189	0.079 to 0.882	0.079 to 0.882	NA
MSIS-8D-P value	1722	-	0.655	0.181	0.138 to 0.893	0.138 to 0.893	NA
ICECAP-A value	1693	-	0.767	0.199	0.071 to 1	0 to 1	0.83
ASCOT value	1670	-	0.771	0.199	−0.062 to 1	−0.171 to 1	0.86
Time point 2							
MSWS-12	1608	-	61.530	37.897	0 to 100	0 to 100	
HADS depression score	1618	-	6.985	4.589	0 to 21	0 to 21	
HADS anxiety score	1618	-	6.682	4.629	0 to 21	0 to 21	
FSS score	1663	-	4.942	1.468	1 to 7	1 to 7	
EQ-5D value	1712	-	0.577	0.305	−0.594 to 1	−0.594 to 1	
MSIS-8D value	1660	-	0.618	0.187	0.079 to 0.882	0.079 to 0.882	
MSIS-8D-P value	1660	-	0.655	0.180	0.138 to 0.893	0.138 to 0.893	
ICECAP-A value	1537	-	0.767	0.197	0 to 1	0 to 1	
ASCOT value	1512	-	0.775	0.201	−0.171 to 1	−0.171 to 1	
Time point 3							
MSWS-12	1366	-	63.472	37.605	0 to 100	0 to 100	
HADS depression score	1367	-	6.822	4.632	0 to 21	0 to 21	
HADS anxiety score	1367	-	6.328	4.736	0 to 21	0 to 21	
FSS score	1205	-	4.989	1.441	1 to 7	1 to 7	
EQ-5D value	1396	-	0.577	0.304	−0.429 to 1	−0.594 to 1	
MSIS-8D value	1382	-	0.622	0.189	0.079 to 0.882	0.079 to 0.882	
MSIS-8D-P value	1382	-	0.660	0.182	0.138 to 0.893	0.138 to 0.893	
ICECAP-A value	1012	-	0.762	0.202	0 to 1	0 to 1	
ASCOT value	1003	-	0.778	0.193	−0.056 to 1	−0.171 to 1	
Spearman's rho	**EQ-5D-3L**		**MSIS-8D**		**MSIS-8D-P**	**ICECAP-A**	**ASCOT**
MSWS-12	−0.584		−0.708		−0.713	−0.568	−0.549
HADS-D	−0.681		−0.771		−0.762	−0.803	−0.724
HADS-A	−0.501		−0.581		−0.570	−0.587	−0.540
FSS	−0.685		−0.650		−0.653	−0.561	−0.519

Note. Sources for UK general population norms: norms are provided for the UK adult general population.
Gender and age: Office for National Statistics: Census-based statistics UK 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/censusbasedstatisticsuk2021

- The figure reported here is the mean average age of people aged 18 or over. Note that people aged 85 years and older were combined into a single category in the published Census-based statistics report. For the purposes of calculating an estimated mean age for this table, the age for this group was assumed to be 90 years.

HADS depression and anxiety scores: Crawford et al.[47]
EQ-5D value: Janssen et al.[48]
ICECAP-A value: Mitchell et al.[49]
ASCOT value: Netten et al.[22]
The MSWS-12, MSIS-8D and MSIS-8D-P are specific to multiple sclerosis, therefore general population values are not relevant. No UK general population norms could be sourced for the FSS.
N indicates number of observations; PPMS, primary progressive multiple sclerosis; RRMS, relapsing remitting multiple sclerosis; SD, standard deviation; SPMS, secondary progressive multiple sclerosis.

Results

Participants

Table 2 presents demographic and clinical data for the 1742 people with MS who provided at least 2 consecutive responses to at least 1 PBM over the 3 time points, as well as mean (SD) PBM values and PROM scores at all time points. At time point 1, mean (SD) health state values were somewhat lower for the EQ-5D-3L (0.567 [0.310]) compared with the MSIS-8D (0.617 [0.189]) and MSIS-8D-P (0.655 [0.181]). Mean capability well-being according to the ICECAP-A value was 0.767 (0.199), and mean SCRQL according to the ASCOT value was 0.771 (0.199).

Table 3. Results of *t* tests, standardized response means and effect sizes, grouped by changes in PROM score.

Time point 1 to 2											
Measure	**Obs**	**Mean (SD) time point 1**	**Mean (SD) time point 2**	**Mean (SD) change score**	**t-stat**	**p-unadj**	**p-adj sig†**	**SRM**	**SES**	**SRM rank**	**Change score rank**
Respondent groups based on change in MSWS-12 scores											
Improvement											
EQ-5D	274	0.590 (0.251)	0.628 (0.240)	0.038 (0.207)	−3.028	0.003	No	0.183	0.154	5	2
MSIS-8D*	275	0.617 (0.168)	0.656 (0.159)	0.039 (0.132)	−4.927	<0.001	**Yes**	0.297	**0.240**	**1**	**1**
MSIS-8D-P*	275	0.657 (0.163)	0.692 (0.155)	0.035 (0.125)	−4.681	<0.001	**Yes**	0.282	**0.222**	2	3
ICECAP-A	260	0.788 (0.182)	0.815 (0.163)	0.027 (0.102)	−4.250	<0.001	**Yes**	0.264	0.156	4	5
ASCOT	250	0.785 (0.173)	0.816 (0.155)	0.032 (0.112)	−4.456	<0.001	**Yes**	0.282	0.193	2	4
Deterioration											
EQ-5D	242	0.585 (0.264)	0.571 (0.260)	−0.014 (0.197)	1.113	0.267	No	−0.072	−0.054	5	5
MSIS-8D	241	0.626 (0.171)	0.601 (0.174)	−0.024 (0.123)	3.079	0.002	No	−0.198	−0.142	2	2
MSIS-8D-P	241	0.668 (0.166)	0.641 (0.172)	−0.028 (0.121)	3.584	<0.001	No	−0.231	−0.165	**1**	**1**
ICECAP-A	230	0.761 (0.207)	0.739 (0.207)	−0.021 (0.111)	2.941	0.004	No	−0.194	−0.104	3	4
ASCOT	223	0.758 (0.213)	0.736 (0.223)	−0.022 (0.115)	2.838	0.005	No	−0.190	−0.101	4	3
Respondent groups based on change in HADS-D scores											
Improvement											
EQ-5D	331	0.479 (0.318)	0.537 (0.309)	0.058 (0.217)	−4.816	<0.001	**Yes**	0.265	0.184	5	2
MSIS-8D*	335	0.543 (0.189)	0.606 (0.179)	0.062 (0.135)	−8.477	<0.001	**Yes**	0.463	**0.339**	**1**	**1**
MSIS-8D-P*	335	0.586 (0.185)	0.644 (0.173)	0.058 (0.132)	−8.09	<0.001	**Yes**	0.442	**0.325**	2	2
ICECAP-A*	305	0.708 (0.205)	0.750 (0.194)	0.042 (0.125)	−5.908	<0.001	**Yes**	0.338	**0.212**	3	5
ASCOT*	300	0.710 (0.205)	0.753 (0.197)	0.043 (0.135)	−5.571	<0.001	**Yes**	0.322	**0.216**	4	4
Deterioration											
EQ-5D	386	0.551 (0.301)	0.518 (0.308)	−0.033 (0.213)	3.05	0.002	No	−0.155	−0.108	5	5
MSIS-8D*	384	0.610 (0.173)	0.562 (0.188)	−0.048 (0.132)	7.123	<0.001	**Yes**	−0.363	**−0.265**	2	2
MSIS-8D-P*	384	0.650 (0.166)	0.600 (0.184)	−0.049 (0.128)	7.558	<0.001	**Yes**	−0.386	**−0.283**	**1**	**1**
ICECAP-A	353	0.756 (0.194)	0.719 (0.204)	−0.037 (0.127)	5.482	<0.001	**Yes**	−0.292	−0.186	3	3
ASCOT	345	0.763 (0.202)	0.729 (0.213)	−0.034 (0.128)	4.9	<0.001	**Yes**	−0.264	−0.162	4	4
Respondent groups based on change in HADS-A scores											
Improvement											
EQ-5D	405	0.508 (0.312)	0.567 (0.295)	0.059 (0.219)	−5.42	<0.001	**Yes**	0.269	0.194	5	**1**
MSIS-8D*	408	0.548 (0.198)	0.605 (0.183)	0.057 (0.134)	−8.617	<0.001	**Yes**	0.427	**0.300**	**1**	2
MSIS-8D-P*	408	0.591 (0.191)	0.643 (0.175)	0.052 (0.130)	−8.005	<0.001	**Yes**	0.396	**0.282**	3	3
ICECAP-A*	375	0.707 (0.214)	0.756 (0.190)	0.049 (0.119)	−8.015	<0.001	**Yes**	0.414	**0.244**	2	4
ASCOT	366	0.719 (0.211)	0.759 (0.201)	0.041 (0.133)	−5.848	<0.001	**Yes**	0.306	0.197	4	5
Deterioration											
EQ-5D	376	0.591 (0.295)	0.539 (0.312)	−0.051 (0.184)	5.432	<0.001	**Yes**	−0.280	−0.170	5	2
MSIS-8D*	378	0.631 (0.176)	0.580 (0.196)	−0.052 (0.131)	7.636	<0.001	**Yes**	−0.393	**−0.277**	2	**1**
MSIS-8D-P*	378	0.669 (0.170)	0.618 (0.192)	−0.051 (0.128)	7.742	<0.001	**Yes**	−0.398	**−0.282**	**1**	2
ICECAP-A*	356	0.768 (0.193)	0.725 (0.209)	−0.044 (0.121)	6.822	<0.001	**Yes**	−0.362	**−0.218**	3	4
ASCOT	346	0.782 (0.190)	0.745 (0.210)	−0.037 (0.123)	5.528	<0.001	**Yes**	−0.297	−0.183	4	5
Respondent groups based on change in FSS scores											
Improvement											
EQ-5D	160	0.585 (0.275)	0.644 (0.245)	0.058 (0.195)	−3.782	<0.001	No	0.299	**0.224**	3	**1**
MSIS-8D*	160	0.621 (0.181)	0.671 (0.164)	0.050 (0.108)	−5.894	<0.001	**Yes**	0.466	**0.290**	**1**	2
MSIS-8D-P*	160	0.658 (0.175)	0.704 (0.161)	0.046 (0.109)	−5.35	<0.001	**Yes**	0.423	**0.275**	2	3
ICECAP-A	147	0.782 (0.186)	0.810 (0.175)	0.028 (0.110)	−3.106	0.002	No	0.256	0.156	4	4
ASCOT	141	0.795 (0.185)	0.812 (0.186)	0.018 (0.128)	−1.631	0.105	No	0.137	0.095	5	5
Deterioration											

continued on next page

Table 3. Continued

Time point 1 to 2											
Measure	**Obs**	**Mean (SD) time point 1**	**Mean (SD) time point 2**	**Mean (SD) change score**	**t-stat**	**p-unadj**	**p-adj sig†**	**SRM**	**SES**	**SRM rank**	**Change score rank**
EQ-5D	160	0.619 (0.300)	0.602 (0.306)	−0.016 (0.195)	1.066	0.288	No	−0.084	−0.054	5	4
MSIS-8D	158	0.666 (0.165)	0.640 (0.176)	−0.026 (0.107)	3.005	0.003	No	−0.239	−0.150	**1**	**1**
MSIS-8D-P	158	0.701 (0.159)	0.677 (0.171)	−0.024 (0.107)	2.86	0.005	No	−0.228	−0.148	2	2
ICECAP-A	143	0.814 (0.176)	0.796 (0.190)	−0.018 (0.107)	1.978	0.050	No	−0.165	−0.097	3	3
ASCOT	135	0.816 (0.163)	0.800 (0.188)	−0.015 (0.118)	1.494	0.138	No	−0.129	−0.086	4	5

Time point 2 to 3											
Measure	**Obs**	**Mean (SD) time point 2**	**Mean (SD) time point 3**	**Mean (SD) change score**	**t-stat**	**p-unadj**	**p-adj sig†**	**SRM**	**SES**	**SRM rank**	**Change score rank**
Respondent groups based on change in MSWS-12 scores											
Improvement											
EQ-5D	170	0.601 (0.241)	0.645 (0.227)	0.043 (0.157)	−3.583	<0.001	No	0.275	0.184	3	**1**
MSIS-8D*	171	0.627 (0.158)	0.666 (0.152)	0.038 (0.112)	−4.471	<0.001	**Yes**	0.342	**0.247**	2	3
MSIS-8D-P*	171	0.661 (0.157)	0.701 (0.149)	0.040 (0.111)	−4.750	<0.001	**Yes**	0.363	**0.264**	**1**	2
ICECAP-A	125	0.781 (0.175)	0.789 (0.189)	0.008 (0.108)	−0.867	0.388	No	0.078	0.046	5	5
ASCOT	124	0.784 (0.173)	0.796 (0.166)	0.012 (0.105)	−1.257	0.211	No	0.113	0.070	4	4
Deterioration											
EQ-5D	224	0.624 (0.236)	0.587 (0.229)	−0.036 (0.196)	2.756	0.006	No	−0.184	−0.155	4	**1**
MSIS-8D*	222	0.661 (0.148)	0.626 (0.158)	−0.035 (0.120)	4.392	<0.001	**Yes**	−0.295	**−0.232**	**1**	2
MSIS-8D-P*	222	0.700 (0.143)	0.666 (0.156)	−0.034 (0.117)	4.344	<0.001	**Yes**	−0.292	**−0.229**	2	3
ICECAP-A	157	0.805 (0.176)	0.787 (0.169)	−0.019 (0.123)	1.892	0.060	No	−0.151	−0.107	5	5
ASCOT	156	0.809 (0.159)	0.777 (0.169)	−0.031 (0.120)	3.264	0.001	No	−0.261	−0.192	3	4
Respondent groups based on change in HADS-D scores											
Improvement											
EQ-5D	292	0.533 (0.316)	0.560 (0.303)	0.027 (0.196)	−2.363	0.019	No	0.138	0.087	5	5
MSIS-8D*	291	0.585 (0.183)	0.630 (0.179)	0.044 (0.112)	−6.76	<0.001	**Yes**	0.396	**0.246**	**1**	**1**
MSIS-8D-P*	291	0.626 (0.179)	0.669 (0.174)	0.043 (0.113)	−6.511	<0.001	**Yes**	0.382	**0.244**	2	2
ICECAP-A	216	0.741 (0.191)	0.770 (0.181)	0.028 (0.125)	−3.354	0.001	No	0.228	0.153	4	4
ASCOT	213	0.754 (0.184)	0.784 (0.178)	0.030 (0.112)	−3.875	<0.001	No	0.265	0.163	3	3
Deterioration											
EQ-5D	273	0.538 (0.291)	0.492 (0.300)	−0.046 (0.208)	3.663	<0.001	No	−0.222	−0.156	5	5
MSIS-8D*	273	0.596 (0.175)	0.544 (0.198)	−0.052 (0.143)	5.973	<0.001	**Yes**	−0.361	**−0.276**	4	3
MSIS-8D-P*	273	0.635 (0.170)	0.583 (0.192)	−0.052 (0.139)	6.201	<0.001	**Yes**	−0.375	**−0.288**	3	3
ICECAP-A*	188	0.752 (0.185)	0.687 (0.211)	−0.065 (0.111)	8.012	<0.001	**Yes**	−0.584	**−0.327**	**1**	**1**
ASCOT*	187	0.774 (0.185)	0.710 (0.214)	−0.064 (0.140)	6.238	<0.001	**Yes**	−0.456	**−0.319**	2	2
Respondent groups based on change in HADS-A scores											
Improvement											
EQ-5D	324	0.570 (0.300)	0.600 (0.290)	0.030 (0.198)	−2.718	0.007	No	0.151	0.101	5	3
MSIS-8D*	325	0.605 (0.192)	0.643 (0.181)	0.038 (0.109)	−6.337	<0.001	**Yes**	0.352	**0.206**	2	**1**
MSIS-8D-P*	325	0.641 (0.185)	0.680 (0.172)	0.038 (0.106)	−6.489	<0.001	**Yes**	0.360	**0.214**	**1**	**1**
ICECAP-A	239	0.763 (0.190)	0.787 (0.191)	0.024 (0.122)	−3.04	0.003	No	0.197	0.126	3	4
ASCOT	238	0.777 (0.185)	0.796 (0.169)	0.018 (0.111)	−2.554	0.011	No	0.166	0.104	4	5
Deterioration											
EQ-5D	289	0.551 (0.297)	0.512 (0.314)	−0.039 (0.196)	3.392	0.001	No	−0.200	−0.128	5	**1**
MSIS-8D	288	0.600 (0.182)	0.562 (0.197)	−0.038 (0.127)	5.068	<0.001	**Yes**	−0.299	−0.199	**1**	2
MSIS-8D-P*	288	0.639 (0.176)	0.602 (0.191)	−0.037 (0.125)	5.021	<0.001	**Yes**	−0.296	**−0.201**	2	3

continued on next page

Table 3. Continued

Time point 2 to 3											
Measure	**Obs**	**Mean (SD) time point 2**	**Mean (SD) time point 3**	**Mean (SD) change score**	**t-stat**	**p-unadj**	**p-adj sig[†]**	**SRM**	**SES**	**SRM rank**	**Change score rank**
ICECAP-A	200	0.751 (0.188)	0.720 (0.192)	−0.031 (0.118)	3.68	<0.001	No	−0.260	−0.162	4	5
ASCOT	198	0.771 (0.196)	0.734 (0.200)	−0.037 (0.143)	3.677	<0.001	No	−0.261	−0.189	3	3
Respondent groups based on change in FSS scores											
Improvement											
EQ-5D	86	0.599 (0.299)	0.635 (0.298)	0.036 (0.234)	−1.434	0.155	No	0.155	0.121	3	3
MSIS-8D	84	0.637 (0.156)	0.682 (0.158)	0.045 (0.114)	−3.661	<0.001	No	0.399	**0.289**	2	**1**
MSIS-8D-P	84	0.673 (0.146)	0.717 (0.150)	0.044 (0.109)	−3.694	<0.001	No	0.403	**0.297**	**1**	2
ICECAP-A	61	0.793 (0.175)	0.794 (0.171)	0.000 (0.119)	−0.027	0.979	No	0.003	0.002	4	4
ASCOT	60	0.815 (0.183)	0.812 (0.175)	−0.003 (0.126)	0.198	0.843	No	−0.026	−0.018	5	5
Deterioration											
EQ-5D	114	0.616 (0.293)	0.562 (0.304)	−0.055 (0.188)	3.094	0.002	No	−0.290	−0.183	4	**1**
MSIS-8D*	114	0.662 (0.159)	0.612 (0.173)	−0.050 (0.131)	4.046	<0.001	**Yes**	−0.379	**−0.300**	**1**	2
MSIS-8D-P	114	0.697 (0.155)	0.648 (0.168)	−0.049 (0.133)	3.908	<0.001	No	−0.366	**−0.302**	2	3
ICECAP-A	83	0.803 (0.163)	0.764 (0.187)	−0.039 (0.107)	3.333	0.001	No	−0.366	**−0.223**	2	4
ASCOT	82	0.803 (0.165)	0.781 (0.182)	−0.022 (0.131)	1.517	0.133	No	−0.168	−0.127	5	5

Note. The following are highlighted in bold: significant results following Bonferroni correction, SES ≥ 0.2, highest ranked SRMs and mean change scores.
Obs indicates observations; p-adj sig, significance following Bonferroni correction; p-unadj, unadjusted *P* value; SD, standard deviation; SES, standardized effect size; SRM, standardized response mean.
*Indicates both statistical significance and SES ≥ 0.2.
[†]Bonferroni correction produced a new threshold for significance of 9.804×10^{-5}.

Responsiveness

Table 3 presents the results of the *t* tests, the SRM and SES, and rankings of SRM and mean change scores, by PBM, for each of the 16 groups (ie, 8 respondent groups in which an effect was hypothesized from time point 1 to 2 and from time point 2 to 3, based on MIDs for the MSWS-12, HADS Depression, HADS Anxiety and FSS). A summary of the responsiveness results for each PBM is given in Table 4.

Bonferroni correction produced a new threshold for significance of 9.804×10^{-5}. Across the 16 groups for which an effect was hypothesized, the number of groups with a statistically significant change in PBM values ranged from 3 for the EQ-5D-3L to 13 for the MSIS-8D. The MSIS-8D-P was responsive to change in 11 groups, and the ICECAP-A and ASCOT were each responsive to change in 6 groups. No statistically significant changes in PBM values were observed in respondents whose PROM scores differed by less than the corresponding MID between time points (results available on request from authors).

In terms of effect sizes, none met the threshold for a moderate or large effect size. Small effects were detected by the EQ-5D-3L in 1 group, the MSIS-8D-P in 13 groups, and the MSIS-8D in 12. The ICECAP-A detected small effects in 5 groups, and the ASCOT in 2.

With regard to SRMs, the EQ-5D-3L had the lowest mean (4.50) and modal (5) average ranked SRM across all the groups and was ranked last (5/5) in 11 groups. The MSIS-8D had the highest mean (1.56) and modal (1) average ranked SRM, with only 1 ranking below 2/5. The mean and modal average rank for the MSIS-8D-P SRMs were 1.75 and 2, respectively, with 2 rankings below 2/5. The ICECAP-A SRMs had a mean and modal average rank of 3.31 and 3, respectively, whereas the SRMs for the ASCOT had a mean rank of 3.75 and mode of 4.

Based on mean change scores, the modal average rank (1) for the EQ-5D-3L was the joint-highest of the 5 PBMs, alongside the MSIS-8D. Its mean average rank (2.63) was the third highest, behind the MSIS-8D (1.69) and MSIS-8D-P (2.25); this was primarily due to the EQ-5D-3L having the smallest mean change score in 4 groups. The MSIS-8D mean change score only ranked below 2/5 in 2 groups. The comparative responsiveness of the MSIS-8D, MSIS-8D-P, ICECAP-A, and ASCOT relative to each other remained consistent with the analyses using standardized values.

Regression Analyses

Table 5 shows the results of the regression analyses in which standardized change scores on each of the PBMs (the dependent variable) were regressed against standardized changes in all PROMs scores (continuous, independent variables). The MSIS-8D measures were the most responsive, showing strong relationships ($P \leq .001$) with all 4 PROMs and standardized beta coefficients ranging from −0.081 to −0.253. ICECAP-A and EQ-5D-3L values were significantly but less strongly related to all PROMs (coefficient ranges: −0.044 to −0.204 and −0.053 to −0.162 respectively). ASCOT values were not significantly related to FSS scores (range of significant coefficients: −0.056 to −0.175).

Table 6 shows the results of the regression analyses in which standardized change scores on each of the PBMs (the dependent variable) were regressed against changes in all PROMs scores categorized according to no change (the baseline level), improvement, or deterioration. Again, the MSIS-8D measures were the most responsive, showing strong or moderate relationships with improvement and deterioration on all PROMs (coefficient range: −0.143 to −0.327). The EQ-5D-3L was the least responsive, showing no significant relationship ($P > .05$) with

Table 4. Summary of responsiveness results.

Measure	Number of groups (total *N* = 16)		Ranked SRM		Ranked mean change score	
	With a statistically significant change	With an SES > 0.2	Mean average rank	Modal average rank	Mean average rank	Modal average rank
EQ-5D-3L	3	1	4.50	5	2.63	1
MSIS-8D	13	12	1.56	1	1.69	1
MSIS-8D-P	11	13	1.75	2	2.25	3
ICECAP-A	6	5	3.31	3	4.00	4
ASCOT	6	2	3.75	4	4.13	5

Note. SRMs and mean change scores were ranked from 1 for the largest absolute value to 5 for the smallest.

improvement or deterioration on the MSWS-12 or deterioration on the FSS, and weaker relationships with improvement or deterioration on HADS-D scores or improvement on the FSS (range of significant coefficients: −0.129 to −0.216).

Illness-Related Events

Results of the IRE analysis are presented in Supplemental Materials (see Tables A1 and A2 in Supplemental Materials). No associations between IREs and PBM values reached significance across time points. Of the 25 mean change scores with SES ≥ 0.2, 13 were in the opposite direction to that hypothesized. Across all the PBMs in the regression analyses, only 5 variables were significant. In all these analyses, results were not consistent across the time points.

Discussion

To our knowledge, this is the first investigation into the responsiveness of PBMs recommended by NICE for use in the economic evaluation of interventions with a social care element, when used with people with MS. The comparison between these and 2 HRQL-PBMs, 1 generic and 1 condition specific, provides evidence to inform the selection of outcome measures for economic evaluations of MS interventions.

Summary of Findings

A clear pattern emerged across the standardized analyses. The EQ-5D-3L was consistently the least responsive of the PBMs. The most responsive were the MSIS-8D and MSIS-8D-P, which performed either best or second-best in the *t* tests, SES and SRM analyses, and equal-best in the regression analyses. The ICECAP-A exhibited slightly better responsiveness than the ASCOT in the SES and SRM analyses but slightly worse in the second regression analyses, and these instruments showed equal responsiveness according to the *t* tests and first regression analysis.

This pattern shifted when absolute, rather than standardized, metrics were used. Here, the EQ-5D-3L appeared substantially more responsive than the SCW-PBMs, performing similarly to the MSIS-8D and MSIS-8D-P. This finding may be an artefact of differences in scale length. EQ-5D-3L values can range from −0.594 to 1,[16] a much greater range of scores than the other PBMs (MSIS-8D: 0.079 to 0.882; MSIS-8D-P: 0.138 to 0.893; ICECAP-A: 0 to 1; ASCOT −0.17 to 1).[19,20,22,23] The standardization inherent in the SRM and SES controls for these differences in scale, whereas the mean change scores do not.

The MS-specific MSIS-8D measures were clearly the most responsive of the PBMs, even when absolute (rather than standardized) mean changes were assessed. This may be unsurprising given that the MSIS-8D descriptive system is condition specific, designed to include content of particular relevance to people

Table 5. Multivariable regression analyses between change in score on the PROMs and the PBMs (*N* = 2226, *n* = 1458).

Measure	EQ-5D-3L	MSIS-8D	MSIS-8D-P	ICECAP-A	ASCOT
MSWS-12	−0.053* (0.020) [−0.093, −0.013]	−0.081*** (0.021) [−0.123, −0.040]	−0.081*** (0.022) [−0.123, −0.038]	−0.053** (0.020) [−0.093, −0.013]	−0.056** (0.020) [−0.095, −0.017]
HADS depression	−0.074** (0.027) [−0.127, −0.022]	−0.187*** (0.026) [−0.238, −0.137]	−0.192*** (0.026) [−0.242, −0.142]	−0.204*** (0.025) [−0.254, −0.155]	−0.175*** (0.024) [−0.223, −0.127]
HADS anxiety	−0.162*** (0.029) [−0.219, −0.105]	−0.253*** (0.026) [−0.304, −0.202]	−0.248*** (0.026) [−0.300, −0.197]	−0.195*** (0.024) [−0.243, −0.147]	−0.170*** (0.024) [−0.217, −0.122]
FSS	−0.063* (0.025) [−0.112, −0.013]	−0.131*** (0.019) [−0.169, −0.094]	−0.132*** (0.019) [−0.170, −0.094]	−0.044* (0.022) [−0.087, 0.000]	−0.044 (0.030) [−0.102, 0.014]
Constant	−0.004 (0.017) [−0.038, 0.030]	0.019 (0.016) [−0.012, 0.051]	0.016 (0.016) [−0.015, 0.048]	0.008 (0.017) [−0.025, 0.040]	0.000 (0.017) [−0.034, 0.034]
R-squared	0.0533	0.1812	0.1811	0.1266	0.0947

Note. Darker shading illustrates a stronger relationship in terms of *P* value.
N indicates number of responses; *n*, number of participants.
*$P \leq .05$, ** $P \leq .01$, *** $P \leq .001$.

Table 6. Multivariable regression analyses between change in score on the PROMs, as indicated by whether an MID occurred indicating an improvement or deterioration in health and change in score on the PBMs (N = 2226, n = 1458).

Measure	*N* (*n*)	EQ-5D-3L	MSIS-8D	MSIS-8D-P	ICECAP-A	ASCOT
MSWS-12	1518 (1072)					
Improvement	349 (346)	0.114 (0.058) [0.000, 0.228]	0.228*** (0.057) [0.117, 0.340]	0.218*** (0.056) [0.109, 0.328]	0.202*** (0.055) [0.093, 0.310]	0.149** (0.053) [0.044, 0.253]
Deterioration	359 (341)	−0.073 (0.056) [−0.182, 0.037]	−0.143** (0.053) [−0.248, −0.039]	−0.159** (0.054) [−0.265, −0.053]	−0.052 (0.053) [−0.155, 0.051]	−0.163** (0.055) [−0.270, −0.056]
HADS-D	1249 (958)					
Improvement	479 (465)	0.129* (0.058) [0.016, 0.242]	0.292*** (0.054) [0.185, 0.398]	0.292*** (0.054) [0.186, 0.399]	0.237*** (0.053) [0.133, 0.340]	0.204*** (0.054) [0.097, 0.310]
Deterioration	498 (483)	−0.129* (0.057) [−0.241, −0.017]	−0.306*** (0.053) [−0.409, −0.202]	−0.330*** (0.053) [−0.433, −0.226]	−0.325*** (0.054) [−0.430, −0.219]	−0.303*** (0.055) [−0.410, −0.196]
HADS-A	1150 (873)					
Improvement	568 (545)	0.195*** (0.055) [0.087, 0.302]	0.327*** (0.050) [0.228, 0.425]	0.310*** (0.050) [0.212, 0.408]	0.348*** (0.051) [0.248, 0.448]	0.237*** (0.050) [0.139, 0.335]
Deterioration	508 (493)	−0.216*** (0.053) [−0.320, −0.111]	−0.278*** (0.051) [−0.379, −0.177]	−0.273*** (0.051) [−0.373, −0.173]	−0.148** (0.051) [−0.248, −0.049]	−0.214*** (0.054) [−0.319, −0.108]
FSS	1810 (1263)					
Improvement	199 (199)	0.201* (0.081) [0.043, 0.360]	0.271*** (0.067) [0.139, 0.403]	0.258*** (0.067) [0.126, 0.390]	0.103 (0.073) [−0.041, 0.247]	−0.004 (0.075) [−0.152, 0.143]
Deterioration	217 (216)	−0.108 (0.071) [−0.247, 0.031]	−0.224*** (0.067) [−0.355, −0.092]	−0.212** (0.068) [−0.346, −0.078]	−0.125 (0.065) [−0.254, 0.003]	−0.071 (0.070) [−0.209, 0.068]
Constant		−0.015 (0.032) [−0.077, 0.047]	−0.007 (0.029) [−0.064, 0.051]	0.003 (0.029) [−0.054, 0.060]	−0.044 (0.028) [−0.098, 0.011]	0.025 (0.031) [−0.037, 0.086]
R-squared		0.0470	0.1482	0.1471	0.1070	0.0839

Note. Darker shading illustrates a stronger relationship in terms of *P* value.
N indicates number of responses; *n*, number of participants.
*$P \leq .05$, ** $P \leq .01$, *** $P \leq .001$.

with MS, whereas the other PBMs are generic. This implies that the relevance of the descriptive system of a PBM can affect its responsiveness. It may also hold that the relevance of the preference weights of a PBM may affect its responsiveness. Qualitative research suggests differences in the rationales of people with MS and the general public for their health state preferences,[50] and significant differences have been identified between QALY weights derived from people with MS and from the general population.[51] However, the comparison between the MSIS-8D, with preference weights elicited from the general population, and the MSIS-8D-P, with preference weights elicited from people with MS, showed no consistent differences between the measures in terms of responsiveness.

Strengths and Limitations

Using data from the UKMSR is a significant strength of this research. The UKMSR is the largest ongoing cohort study of people with MS in the United Kingdom. It has been described as a representative population of people with MS,[52] and the data have been used in a variety of research studies.

There are limitations to the analysis described in this article. First, at the time of data collection, the UKMSR was using the EQ-5D-3L, rather than the EQ-5D-5L. The increased number of response levels per dimension on the latter is intended to increase its sensitivity to differences and responsiveness to changes in HRQL.[53] The UKMSR has since replaced the EQ-5D-3L with the EQ-5D-5L, and future analyses of UKMSR data could usefully explore whether this new version of the instrument demonstrates improved responsiveness in the context of MS.

The lack of any consistent pattern of responsiveness of the PBMs to MS illness-related events may be due to limitations in how the events were described, defined, and/or captured. However, great methodological care was taken in developing the IRE questionnaire including drawing on the relevant literature, multiple patient involvement sessions, cognitive interviews, and piloting via the UKMSR. Limited responsiveness of PBMs to key events that impact the quality of life and well-being of people with MS is of concern.

Consequently, the responsiveness of the PBMs discussed here is based on changes in PROMs that assess symptoms commonly experienced by people with MS. The constructs measured by these PROMs (walking ability, depression, anxiety, and fatigue) are of greater relevance to the HRQL-PBMs than to the SCW-PBMs and as such may underestimate the responsiveness of the latter, particularly given that SCW-PBMs are recommended for use in social care interventions that are unlikely to affect symptoms directly.

It is not yet known what drives the differences in responsiveness of the PBMs.[54] The conceptualization of HRQL used in relation to PBMs differs from its broader conceptualization in health and quality-of-life research. HRQL in a wider context has been defined as, "the patient's subjective perception of the impact of his [*sic*] disease and its treatment(s) on his [*sic*] daily life, physical, psychological and social functioning and well-being."[55] In the PBM framework, HRQL is operationalized as the preference weight assigned to an individual's health state based (usually) on general population preferences for that health state. A change in a HRQL-PBM value indicates a change in health status, represented by the associated change in preference weight. Likewise, for the SCW-PBMs, the construct being considered is the preference weight assigned to an individual's "well-being state" or "SCRQL state." A change in a SCW-PBM value indicates a change in well-being or SCRQL state, represented by the associated change in preference weight. Therefore, the descriptive systems of the PBMs may be

responsive to change, but it is ultimately the overlying preference weights associated with these underlying states that will dictate the responsiveness of the measure. These components of PBMs and how to investigate their psychometric properties require further consideration and exploration.

Conclusions

The relative responsiveness of SCW-PBMs versus HRQL-PBMs in the context of MS appears to depend on which PBMs are being considered and the metric(s) used to judge responsiveness.

These findings indicate that a condition-specific HRQL-PBM was more responsive to changes in walking ability, anxiety, depression, and fatigue compared with the SCW-PBMs. Compared with a generic HRQL-PBM, the SCW-PBMs demonstrated greater responsiveness.

However, this is only true if the responsiveness of the measures is compared using standardized scores; the wide range of EQ-5D-3L tariff values without standardization tends toward eclipsing the apparent responsiveness of other PBMs. This is an important consideration in the selection of PBMs for use in economic evaluation: it is the magnitude of change in absolute values that determines the effectiveness side of the cost-effectiveness equation; therefore, the scale-range of a PBM could influence whether an intervention is found to be cost-effective. NICE currently recommend the use of SCW-PBMs as a complement to HRQL-PBMs to capture nonhealth effects of interventions.[15] These findings suggest that SCW-PBM analyses require careful interpretation and presentation to ensure that nonhealth impacts are not eclipsed by "larger" health impacts when the 2 are used together to inform health and social care decision making.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2025.09.3063.

Article and Author Information

Accepted for Publication: September 26, 2025

Published Online: November 15, 2025

doi: https://doi.org/10.1016/j.jval.2025.09.3063

Author Affiliations: University of Exeter Medical School, University of Exeter, Exeter, England, UK (Goodwin, Boddy, Hawton); PenCHORD (The Peninsula Collaboration for Health Operational Research and Data Science), Department of Health and Community Sciences, University of Exeter, Exeter, England, UK (Heather); Public Health and Sport Sciences, University of Exeter, Exeter, England, UK (Morrish); School of Health Professions, Faculty of Health, University of Plymouth, Plymouth, England, UK (Freeman); Faculty of Health and Social Sciences, University of Bournemouth, Bournemouth, England, UK (Thomas); UCL Queen Square Institute of Neurology, University College London, London, England, UK (Chataway); Population Data Science, Swansea University, Swansea, Wales, UK (Middleton).

Correspondence: Elizabeth Goodwin PhD, University of Exeter Medical School, St Luke's Campus, Exeter, EX1 2LU, United Kingdom. Email: e.goodwin@exeter.ac.uk

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: This work was supported by Grant 82 from the UK Multiple Sclerosis Society.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Availability: Materials used in this study can be found at https://github.com/exeterhems/hums_responsiveness and https://doi.org/10.5281/zenodo.15095009.

Acknowledgment: The research reported in this article was funded by the Multiple Sclerosis Society of Great Britain (Grant Code 82). The data were collected by the UK MS Register Project (Ref: 16/SW/0194); the authors acknowledge the contribution of all Members of the UK MS Register Team based at Swansea University Medical School. The authors also thank the Members of the HEMS Patient Involvement Group, and the Members of the MS Register who provided the data used in this study.

REFERENCES

1. Dakin H, Devlin N, Feng Y, Rice N, O'Neill P, Parkin D. The influence of cost-effectiveness and other factors on NICE decisions. *Health Econ*. 2015;24(10):1256–1271.
2. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. 2nd ed. Oxford, United Kingdom: Oxford University Press; 2016.
3. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 2016;316(10):1093–1103.
4. Fisk JD, Brown MG, Sketris IS, Metz LM, Murray TJ, Stadnyk KJ. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *J Neurol Neurosurg Psychiatry*. 2005;76(1):58–63.
5. Hawton A, Green C. Health utilities for multiple sclerosis. *Value Health*. 2016;19(4):460–468.
6. Hawton A, Goodwin E, Boddy K, et al. Measuring the cost-effectiveness of treatments for people with multiple sclerosis: beyond quality-adjusted life-years. *Mult Scler*. 2022;28(3):346–351.
7. McCabe C, Chilcott J, Claxton K, et al. Continuing the multiple sclerosis risk sharing scheme is unjustified. *BMJ*. 2010;340:c1786.
8. Phillips CJ, Humphreys I. Assessing cost-effectiveness in the management of multiple sclerosis. *Clinicoecon Outcomes Res*. 2009;1:61–78.
9. Sharac J, McCrone P, Sabes-Figuera R. Pharmacoeconomic considerations in the treatment of multiple sclerosis. *Drugs*. 2010;70(13):1677–1691.
10. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol*. 2001;54(12):1204–1217.
11. NICE. *Beta Interferon and Glatiramer Acetate for the Treatment of Multiple Sclerosis*. London: National Institute for Clinical Excellence; 2002.
12. NICE. *Natalizumab for the Treatment of Adults With Highly Active Relapsing-Remitting Multiple Sclerosis*. London: National Institute for Clinical Excellence; 2007. https://www.nice.org.uk/guidance/ta127. Accessed October 11, 2025.
13. NICE. *Fingolimod for the Treatment of Highly Active Relapsing-Remitting Multiple Sclerosis*. London: National Institute for Clinical Excellence; 2012. https://www.nice.org.uk/guidance/ta254. Accessed October 11, 2025.
14. Brazier J, Tsuchiya A. Improving cross-sector comparisons: going beyond the health-related QALY. *Appl Health Econ Health Policy*. 2015;13(6):557–565.
15. NICE. *Developing NICE Guidelines: the Manual (PMG20). NICE Process and Methods*. London: National Institute for Health and Care Excellence; 2024. https://www.nice.org.uk/process/pmg20. Accessed October 11, 2025.
16. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
17. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;13(12):1–177. iii, ix-x.
18. Goodwin E, Green C. A quality-adjusted life-year measure for multiple sclerosis: developing a patient-reported health state classification system for a multiple sclerosis-specific preference-based measure. *Value Health*. 2015;18(8):1016–1024.
19. Goodwin E, Green C, Spencer A. Estimating a preference-based index for an eight-dimensional health state classification system for multiple sclerosis. *Value Health*. 2015;18(8):1025–1036.
20. Goodwin E, Green C, Hawton A. Health state values derived from people with multiple sclerosis for a condition-specific preference-based measure: multiple sclerosis Impact Scale-Eight Dimensions-Patient Version (MSIS-8D-P). *Value Health*. 2018;21(11):1338–1345.

21. Engel L, Bryan S, Whitehurst DGT. Conceptualising 'benefits beyond health' in the context of the quality-adjusted life-year: a critical interpretive synthesis. *Pharmacoeconomics*. 2021;39(12):1383–1395.
22. Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol Assess*. 2012;16 (16):1–166.
23. Flynn TN, Huynh E, Peters TJ, et al. Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health Econ*. 2015;24 (3):258–269.
24. Verkerk MA, Busschbach JJ, Karssing ED. Health-related quality of life research and the capability approach of Amartya Sen. *Qual Life Res*. 2001;10 (1):49–55. ; discussion 57-58.
25. Eccles FJ, Simpson J. A review of the demographic, clinical and psychosocial correlates of perceived control in three chronic motor illnesses. *Disabil Rehabil*. 2011;33(13-14):1065–1088.
26. Gruenewald DA, Higginson IJ, Vivat B, Edmonds P, Burman RE. Quality of life measures for the palliative care of people severely affected by multiple sclerosis: a systematic review. *Mult Scler*. 2004;10(6):690–704.
27. Mitchell AJ, Benito-Leon J, Gonzalez JM, Rivera-Navarro J. Quality of life and its assessment in multiple sclerosis: integrating physical and psychological components of wellbeing. *Lancet Neurol*. 2005;4(9):556–566.
28. Nortvedt MW, Riise T. The use of quality of life measures in multiple sclerosis research. *Mult Scler*. 2003;9(1):63–72.
29. Bass AD, Van Wijmeersch B, Mayer L, et al. Effect of multiple sclerosis on daily activities, emotional well-being, and relationships: the global vsMS survey. *Int J MS Care*. 2020;22(4):158–164.
30. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Measuring the impact of MS on walking ability: the 12-Item MS Walking Scale (MSWS-12). *Neurology*. 2003;60(1):31–36.
31. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361–370.
32. Flachenecker P, Kumpfel T, Kallmann B, et al. Fatigue in multiple sclerosis: a comparison of different rating scales and correlation to clinical parameters. *Mult Scler*. 2002;8(6):523–526.
33. Mehta L, McNeill M, Hobart J, et al. Identifying an important change estimate for the multiple sclerosis Walking Scale-12 (MSWS-12v1) for interpreting clinical trial results. *Mult Scler J Exp Transl Clin*. 2015;1: 2055217315596993.
34. Learmonth YC, Dlugonski DD, Pilutti LA, Sandroff BM, Motl RW. The reliability, precision and clinically meaningful change of walking assessments in multiple sclerosis. *Mult Scler*. 2013;19(13):1784–1791.
35. Hobart J, Ziemssen T, Feys P, et al. Assessment of clinically meaningful improvements in self-reported walking ability in participants with multiple sclerosis: results from the randomized, double-blind, phase III ENHANCE trial of prolonged-release fampridine. *CNS Drugs*. 2019;33(1):61–79.
36. Chan KS, Aronson Friedman L, Bienvenu OJ, et al. Distribution-based estimates of minimal important difference for hospital anxiety and depression scale and impact of event scale-revised in survivors of acute respiratory failure. *Gen Hosp Psychiatry*. 2016;42:32–35.
37. Lemay KR, Tulloch HE, Pipe AL, Reed JL. Establishing the minimal clinically important difference for the Hospital Anxiety and Depression Scale in patients with cardiovascular disease. *J Cardiopulm Rehabil Prev*. 2019;39(6):E6–E11.
38. Puhan MA, Frey M, Buchi S, Schünemann HJ. The minimal important difference of the Hospital Anxiety and Depression Scale in patients with chronic obstructive pulmonary disease. *Health Qual Life Outcomes*. 2008;6:46.
39. Robinson Jr D, Zhao N, Gathany T, Kim LL, Cella D, Revicki D. Health perceptions and clinical characteristics of relapsing-remitting multiple sclerosis patients: baseline data from an international clinical trial. *Curr Med Res Opin*. 2009;25(5):1121–1130.
40. Rooney S, McFadyen DA, Wood DL, Moffat DF, Paul PL. Minimally important difference of the fatigue severity scale and modified fatigue impact scale in people with multiple sclerosis. *Mult Scler Relat Disord*. 2019;35:158–163.
41. Ford DV, Jones KH, Middleton RM, et al. The feasibility of collecting information from people with multiple sclerosis for the UK MS Register via a web portal: characterising a cohort of people with MS. *BMC Med Inform Decis Mak*. 2012;12:73.
42. Goodwin E, Heather A, Morrish N, et al. HUMS responsiveness. Zenodo. https://doi.org/10.5281/zenodo.15095009; 2025. Accessed October 11, 2025.
43. Cohen J. *Statistical Power Analysis for Behavioural Sciences*. Hillsdale, NJ: Erlbaum; 1988.
44. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res*. 2003;12(4):349–362.
45. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol*. 2000;53(5):459–468.
46. Hawton A, Boddy K, Kandiyali R, Tatnell L, Gibson A, Goodwin E. Involving patients in health economics research: "The PACTS Principles". *Patient*. 2021;14(4):429–434.
47. Crawford JR, Henry JD, Crombie C, Taylor EP. Brief report: Normative data for the HADS from a large non-clinical sample. *British J Clin Psychol*. 2001;40:429–434.
48. Janssen B., Szende A. Population Norms for the EQ-5D. 2013 Sep 26. In: Szende A., Janssen B., Cabases J., editors. Self-Reported Population Health: An International Perspective based on EQ-5D. Dordrecht, NL: Springer; 2014.
49. Mitchell PM, Al-Janabi H, Byford S, et al. Assessing the validity of the ICECAP-A capability measure for adults with depression. *BMC Psychiatry*. 2017;17:46.
50. Goodwin E, Davey A, Green C, Hawton A. What drives differences in preferences for health states between patients and the public? A qualitative investigation of respondents' thought processes. *Soc Sci Med*. 2021;282: 114150.
51. Goodwin E, Green C, Hawton A. What difference does it make? A comparison of health state preferences elicited from the general population and from people with multiple sclerosis. *Value Health*. 2020;23(2):242–250.
52. Middleton RM, Rodgers WJ, Chataway J, et al. Validating the portal population of the United Kingdom multiple sclerosis Register. *Mult Scler Relat Disord*. 2018;24:3–10.
53. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20 (10):1727–1736.
54. Rowen D, Young T, Brazier J, Gaugris S. Comparison of generic, condition-specific, and mapped health state utility values for multiple myeloma cancer. *Value Health*. 2012;15(8):1059–1068.
55. Riazi A. Patient-reported outcome measures in multiple sclerosis. *Int MS J*. 2006;13(3):92–99.