Patient-reported outcome measures in Multiple Sclerosis

A Riazi

Address for correspondence:

Dr. Afsane Riazi, Psychology Department, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK
Tel: 01784 443706
Fax: 01784 434347
E-mail: Afsane.Riazi@rhul.ac.uk

Summary

Patient reported outcome (PRO) measures are increasingly being used in epidemiological studies, health services research and in clinical trials to evaluate therapeutic interventions from the patient's perspective. In MS, where evaluations of treatments are becoming increasingly important, outcome measures that are rigorous and appropriate to patients are needed. This article examines the value of PRO measures in MS, the criteria of evaluating such measures, and provides a brief description of MS-specific PRO measures that are currently available.

Introduction

Patient reported outcome (PRO) is an umbrella term referring to questionnaires, interviews and other related methods of assessing health, illness and benefits of health care interventions from the patient's perspective. For people with chronic illnesses such as MS, these measures capture the diverse impact of illnesses on their day-to-day lives.

Over the past two decades, outcome measurement in MS has relied extensively on the Expanded Disability Status Scale (EDSS) [1] as a measure of disease activity. However, the EDSS is heavily weighted towards mobility and does not reflect patients' problems in other areas of health. Furthermore, the EDSS was developed before psychometric methods became familiar to clinicians, was not based on recognised techniques of scale constructions [2], and has limited measurement properties [3,4]. Most importantly, it is rated by neurolgists and may not adequately reflect patients' own perceptions of the impact of their MS.

This paper aims to highlight: 1) the importance of PRO measures 2) criteria for selecting PRO measures and 3) a brief review of currently available PRO measures in MS.

What are PRO measures and why are they important?

A simple but useful classification considers health outcomes to be either *physician* or *patient-based.*

Physician-based outcomes are the traditional biomedically defined outcomes. In MS, MRI and relapse rate are good examples of physician-based outcomes. Although these assessments are in no doubt important, they only partly address partly address patients' concerns [5], and provide an incomplete picture of the disease impact of MS from the patients' perspectives.

Patient-based, or patient-reported outcomes (PROs), on the other hand, are concerned with the consequences of disease and treatment that are considered important to patients. As health care providers place increasing emphasis on evidence-based practice, PRO measures have become increasingly important, as they measure outcomes that matter to patients. There is evidence that MS patients and physicians differ in which domains of health are the most important [6]. As new treatments for MS are aimed at altering its natural history or modifying

its impact, PRO measures are essential in a comprehensive evaluation of therapeutic effectiveness.

Although all PRO instruments address some aspect of the patient's subjective experience of their health, there is lack of concensus with regards to what PROs actually measure. For example, there is little consistency of use or agreement in the meaning of terms such as 'quality of life', 'health-related quality of life', 'health status', 'functional status' and 'functional well-being'. However, patient-based outcome measures most often focus on what is commonly referred to as “health-related quality of life” (HRQoL).

Health-related quality of life (HRQoL)

Although many definitions for HRQoL have been proposed, the most accepted definition is *“the patient’s subjective perception of the impact of his disease and its treatment(s) on his daily life, physical, psychological and social functioning and well-being”* [7]. This definition of HRQL has as a common basis with the definition of health given by the WHO in 1948: *“Health, is a state of complete physical, mental, and social well-being and not merely the absence of disease”*[8]. Furthermore, it is generally accepted that while “quality of life (QoL)” is to do with a much broader concept, including community, work and family domains [9], “HRQoL” concerns those domains that are most affected by disease, injury or treatment [10].

Several characteristics of HRQoL have also been proposed. These include: *multidimensionality*, *subjectivity*, and *self-administration*.

First, HRQoL is assumed to consist of several dimensions. Again, although there are no consensus as to the number and nature of the dimensions, it is generally assumed that HRQoL consists of at least the physical and psychological dimensions [11]. This distinction is empirically supported by many studies, for example, the analyses of the Medical Outcomes Short-form Health Survey (SF-36) [12]. More recently, in addition to the physical and psychological dimensions, the inclusion of the participation (role and social) dimension has been proposed [11].

The second and third characteristics are closely linked. HRQoL is essentially a subjective concept that must be evaluated from the perspective of the patient. The final characteristic is self-administration. Because HRQoL is subjective, patients themselves are most suited to

complete the measures themselves, and there is concern that external administration may somehow influence patient's true responses.

Are there different types of patient-based outcome measures?

Broadly speaking, there are two types of patient-based outcome measures:

Generic measures are those that are designed to be broadly applicable across different types and severity of disease, medical interventions, and demographic and cultural groups so as to permit comparisons across studies. The Medical Outcome Study Short From-36 [13] is a well-known example, and one that has also been used in many studies involving MS patients. For example, in one study, MS patients and Parkinson's disease patients had significantly worse health than the general population on all eight domains measured by the SF-36 [14].

Disease-specific measures are designed to reflect clinically relevant issues for a specific disease. They are intended to have very relevant content, with the items in the questionnaire being developed specifically to assess the impact of the particular disease. Several disease-specific measures for MS patients have now been developed. These are examined in a later section.

Although generic measures have the advantage of enabling comparisons across diseases, it is increasingly recognised that they do not cover some areas of outcome that are highly relevant in specific diseases [15], and may have limited responsiveness [16]. Furthermore, the SF-36 has psychometric limitations when used in MS. These include significant floor and ceiling effects [17], limited responsiveness [17], underestimation of mental health problems [18], and a failure to satisfy assumptions for generating summary scores [19].

How can you evaluate patient-based outcome measures?

Guidance on the evaluation of PRO instruments has been offered by the Division of Drug Marketing, Advertising, and Communications (DDMAC) of the Federal Drug Agency **(**FDA), as well as being available recently in a draft guidance document from the FDA [20]. According to the draft guidance document, PRO instrument development and modification process include the establishment of conceptual framework and identification of the intended application, instrument development, the assessment of measurement properties and instrument modification issues. In particular, for potential users of an PRO measures, the

measurement properties are particularly important in the selection of measures. These measurement properties are based on *psychometric theory*, which is a scientifically rigorous field that is concerned with the science of assessing the measurement characteristics of instruments. In general, there are six psychometric properties that should be examined: *data quality, scaling assumptions, acceptability, reliability, validity and responsiveness*.

Insert Table 1 here

- Indicators of *data quality* such as item non-response and missing scale scores, determine the extent to which an instrument can be used successfully in a clinical setting.
- *Scaling assumptions* test whether items are correctly grouped into scales, and if the items can be summed without weighting or standardisation to produce a score.
- *Acceptability* is concerned with the score distribution of the scale, and whether it represents the true distribution of the construct being measured in the sample.
  The *reliability* of an instrument is defined as the extent to which it is free from random error. A reliable measure produces results that are accurate, consistent, stable over time and reproducible. *Test-retest reliability* (stability of scores over time when no change has occurred in the concept of interest) and *internal consistency* (the intercorrelations of items in the same domain, measured by the internal consistency statistic, e.g. the Cronbach's alpha) are two of the most frequently examined types of reliability.
- *Validity* can be broadly defined as the extent to which an instrument measures the concept it purports or is intended to measure. *Face validity (*the extent to which a measure appears on the surface to measure what it is suppose to measure), *content-related validity* (whether items and response options are relevant and are comprehensive measures of the domain or concept), *construct-related validity* (whether relationships among items, domains, and concepts conform to what is predicted by the conceptual framework for the PRO instrument itself and its validation hypotheses) and *predictive validity* (whether items and response options are relevant and are comprehensive measures of the domain or concept) are frequently examined.
- *Responsiveness* is the ability of an instrument to measure clinically important change over time, and is essential when evaluating the relative benefits of different interventions.

What measures have been developed for MS?

A number of MS-specific measures have been developed. Those measures whose original version is in English include: the Multiple Sclerosis Quality of Life-54 (MSQOL-54) [21], the Functional Assessment of MS (FAMS) [22], the MS Quality of Life Inventory (MSQLI) [23], the Multiple Sclerosis Impact Scale (MSIS-29) [24] and the Leeds Multiple Sclerosis Quality of Life (Leeds MSQoL) [25]. Table 2 reports the summary characteristics of these measures. Brief descriptions of each of the measures are also reported below.

Insert Table 2 here

Multiple Sclerosis Quality of Life-54 (MSQOL-54)

MSQOL-54 is a MS-specific quality of life measure consisting of 54 items. This measure is based on the SF-36, but supplemented with 18 additional items in the following areas: health distress (four items), sexual function (four items), satisfaction with sexual function (one item), overall quality of life (two items), cognitive function (four items), energy (one item), pain (one item) and social function (one item).

Functional Assessment in MS (FAMS)

Cella et al (1996) developed a quality of life instrument consisting of 28 items from the general version of the Functional Assessment of Caner Therapy quality of life instrument, plus 60 items generated by patients, care providers and literature review. Using principal components analyses and Rasch analyses, items were reduced to 44 with subscales: mobility, symptoms, emotional well-being (depression), general contentment, thinking/fatigue, and family/social well-being. Fifteen initially rejected questions were then added back as miscellaneous (unscored) questions for their potential clinical and empirical value, resulting in a final 59-item questionnaire.

MS Quality of Life Inventory (MSQLI)

The MSQLI is a battery consisting of 10 individual scales (SF-36, Modified Fatigue Impact Scale (MFIS); MOS Pain Effects Scale (PES), Sexual Satisfaction Scale (SSS), Bladder

Control Scale (BLCS), Bowel Control Scale (BWCS), Impact of Visual Impairment Scale (IVIS), Perceived Deficits Questionnaire (PDQ). This provides a quality of life measure that is both generic and MS-specific [26]. Some scales also have a short-form. Each of the individual scales generates a separate score, and there is no global composite combining all the scales into a single score.

Multiple Sclerosis Impact Scale (MSIS-29)

The Multiple Sclerosis Impact Scale (MSIS-29) is a measure of the physical and psychological impact of MS from the patient's perspective. The total scale consists of 29 items, while the physical scale consists of 20 items, and the psychological scale consists of 9 items. The scale was developed from in-depth interviews of a community sample of people with MS.

The Leeds Multiple Sclerosis Quality of Life (Leeds MSQoL)

Leeds MSQOL is an eight-item scale developed from focus group sessions with people with MS. Twenty-five initial items were reduced using traditional psychometric methods and Rasch measurement model. The instrument is brief and measures a construct related to well-being.

Other scales developed for people with MS include the RAYS Scale [28] and the Disability and Impact Profile (DIP) [29]. The RAYS Scale is a 50-item questionnaire measuring similar dimensions to MSQLI, MQOL-54 and FAMS. The DIP is a 39 item measure in which patients indicate both their level of disability and the importance of the disability to the patient.

How can you select which instrument to use?

It can be difficult for health professionals to choose between the various MS-specific scales that are available. It may be tempting to choose a measure based on practical reasons, for example, selecting a measure because it is short with few items. However, it is recommended that the choice of patient-based outcome measures be evidence-based; that is, how well the

measure meets the psychometric criteria discussed above. One method is to conduct a head-to-head comparison of various measures and selecting the measure with the best psychometric properties. There are now several studies that have attempted to provide such data (eg. [30, 31]).

Several other review articles on PRO measures in MS have now also been published [eg. 32,33,34] some of which have compared the measurement properties of various measures. For example, Nicholl [32], compared the measurement properties of several measures used in MS and concluded that FAMS was the most superior. These reviews also highlight the fact that PRO measures are useful but that there is a lack of data on comparison of responsiveness of measures [33], and that further work is required to decide which scale is most suited to which purpose [34].

Of course, that is not to say that a brief, user-friendly, and cost effective, measure will be more suitable for use in clinical practice. But in reality, the use of patient-based outcome measures in clinical practice for people with MS is still limited. In a recent review [35], the cultural, practical and methodological reasons behind this were discussed. These included clinicians' lack of knowledge of patient-based outcomes, and the logistic and financial implications of administering, processing and scoring the measures. An additional important point worth mentioning is the fact that currently, the confidence intervals around most patient-based outcome scores are too wide for reliable and valid individual patient clinical decision-making. Newer psychometric approaches, such as Rasch analysis [36], offer the ability to construct interval-level measurements from ordinal-level scales (that are currently derived from existing patient-based outcome measures data). There is now at least one measure for use with people with MS that has been developed guided by Rasch analysis [37]. The use of such newer psychometric approaches are also exciting from the point of view that it can form a basis of computerised administration of measures (computer adaptive testing) [38].

Conclusion

In conclusion, PRO measures are increasingly recognised as being central in health-care evaluation as they offer patients' perspectives on the quality of their health. It is important that such instruments are scientifically rigorous to ensure that interventions for MS are accurately evaluated. This is imperative as an increasing number of clinical trials are being conducted to investigate novel therapeutic strategies for MS [39]. Comprehensive evaluations of the psychometric properties of available measures, ideally in comparison with each other, will

assist in the selection of the most appropriate measure. Although the use of PRO is still not widespread within clinical practice, this may change overtime, as clinicians become more familiar with the importance of obtaining patient-derived data, and newer psychometric approaches offer potentially new and simpler ways of administering the measures (eg use of computer adaptive testing). This will also enable PRO measures to be utilised for individual patient monitoring within routine clinical practice.

References

1. Kurtzke JF. Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983; 33: 1444-52.
2. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.
3. Sharrack B, Hughes RAC, Soudain S, Dunn G. The psychometric properties of clinical rating scales used in multiple sclerosis. Brain 1999;122:141-159.
4. Hobart JC, Freeman JA, Thompson AJ. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. Brain 2000;123:1027-1040.
5. Guyatt GH, Freeny DH, Patrick DL. Measuring health-related quality of life. Annals of Internal Medicine 1993;118(8):622-629.
6. Rothwell PM, McDowell Z, Wong CK, Dorman PJ. Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. BMJ 1997; 314:1580.
7. European Medicines Agency, Committee for Medicinal Products for Human Use. Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products. 2005: London, UK.
8. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.
9. Campbell A, Converse PE, Rodgers WL. The quality of American life. New York: Sage Foundation; 1976.
10. Patrick DL, Erickson P. Health status and health policy: allocating resources to health care. New York: Oxford Univ Pr; 1993.
11. Ware JE Jr. Conceptualization and measurement of health-related quality of life: comments on an evolving field.Arch Phys Med Rehabil 2003;84 Suppl 2:S43-51.
12. Essink-Bot ML, Krabbe PF, Bonsel GJ, Aaronson NK. An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. Med Care 1997;35:522-37.
13. Ware JE Jr, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey manual and interpretation guide. Boston, Massachusetts: Nimrod Press; 1993.

14. Riazi A, Hobart JC, Lamping DL, et al. Using the SF-36 measure to compare the health impact of multiple sclerosis and Parkinson's disease with normal population health profiles. J Neurol Neurosurg Psychiatry. 2003;74:710-714.
15. Peto V, Jenkinson C, Fitzpatrick R, Greenhall R. The development and validation of a short measure of functioning and well-being for individuals with Parkinson's Disease. Quality of Life Research 1995;4:241-248.
16. Patrick D, Deyo R. Generic and disease-specific measures in assessing health status and quality of life. Medical Care 1989;27(3 Suppl):S217-S232.
17. Freeman JA, Hobart JC, Langdon DW, Thompson AJ. Clinical appropriateness: a key factor in outcome measure selection. The 36-item Short Form Health Survey in multiple sclerosis. Journal of Neurology, Neurosurgery, and Psychiatry 2000;68:150-156.
18. Norvedt MW, Riise T, Myer K-M, Nyland HI. Performance of the SF-36, SF-12 and RAND-36 summary scales in a multiple sclerosis population. Medical Care 2000;38:1022-1028.
19. Hobart JC, Freeman JA, Lamping DL, Fitzpatrick R, Thompson AJ. The SF-36 in MS: why assumptions must be tested (abstract). Annals of Neurology 2000;48:495.
20. US Department of Health and Human Services. Draft guidance for industry: Patient-reported outcome measures: use in medical product development to support labelling claims. 2006.
21. Vickrey BG, Hays RD, Harooni R, Myers LW, Ellison GW. A health-related quality of life measure for multiple sclerosis. Quality of life research 1995;4:187-206.
22. Cella DF, Dineen K, Arnason B, Reder A, Webster KA, Karabatsos G, et al., Validation of the Funtional Assessment of Multiple Sclerosis quality of life instrument. Neurology 1996; 47:129-139.
23. Ritvo PG, Fischer JS, Miller DM, et al. Multiple Sclerosis Quality of Life Inventory: A User's Manual. New York: National Multiple Sclerosis Society, 1997.
24. Hobart JC, Lamping DL, Fitzpatrick A, Thompson AJ. The Muliple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. Brain 2001, 124: 962-973.
25. Ford HL, Gerry E, Tennant A, Whalley D, Haigh, R, Johnson MH. Developing a disease-specific quality of life measure for people with multiple sclerosis. Clinical Rehabilitation 2001; 15 (3) 247-58.
26. Fischer JS, LaRocca NG, Miller DM, et al. Recent developments in the assessment of quality of life in multiple sclerosis. Mult Scler. 1999;5:251-259.
27. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome. Health Technol Asesss 2004; 8(9).

28. Rotsstein Z, Barak Y, Noy S, Achiron A. Quality of life in multiple sclerosis: development and validation of the 'RAYS' Scale and comparison with the SF-36. Int J Qual Health Care 2000; 12 (6), 511-517.
29. Lankhorst GJ, Jelles F, Smits RCF et al. Quality of life in multiple slcerosis: the disability and impact profile (DIP). J Neurol 1996; 243 (6), 469-474.
30. Riazi A, Hobart JC, Lamping DL, Fitzpatrick R, Thompson AJ. Evidence-based measurement in multiple sclerosis: the psychometric properties of the physical and psychological dimensions of three quality of life rating scales. Multiple Sclerosis 2003; 9: 411-419
31. Hobart, JC, Riazi, A, Lamping, DL, Fitzpatrick, R, Thompson, AJ. 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.
32. Nicholl CR, et al Assessing quality of life in people with multiple sclerosis Disability and rehabilitation 2001;23(14): 597-603.
33. Duff SB, Mordin MM. Strategies for assessing health economics and quality of life in multiple sclerosis. Expert Rev Pharmacoeconomics and Outcomes Research 2002;2:577-587.
34. Benito-Leon J, Morales JM, Rivera-Navarro J, Mitchell A. A review about the impact of multiple sclerosis on health-related quality of life. Disabil Rehabil 2003 Dec 2; 25 (23): 1291-303.
35. Solari A. Role of health-related quality of life measures in the routine care of people with multiple sclerosis. Health and Quality of Life Outcomes 2005, 3:16
36. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Education Research; 1960. Reprinted Chicago: University of Chicago Press; 1980
37. Hobart JC, Riazi A, Thompson AJ, Styles IM, Ingram W, Vickery PM, Warner M, Fox PJ, and Zajicek JP. Getting the measure of spasticity in multiple sclerosis: the Multiple Sclerosis Spasticity Scale (MSSS-88). Brain 2006 129: 224-234.
38. Wainer H. Computer adaptive testing: A primer. 1990; Hillsdale, NJ: Lawrence Erlbaum.
39. Virley DJ. Developing therapeutics for the treatment of multiple sclerosis. NeuroRX 2005 2(4): 638-49.

Table 1. Examples of psychometric criteria often used for determining the adequacy of HRQoL scales

Psychometric property	**Criterion for adequacy**
Data quality	Missing item data < 10% High % computable scale scores
Scaling assumptions	Similar response option frequency distributions Similar mean scores and variances
Acceptability	Scores span the full scale range Mean scores near midpoint
Reliability	Cronbach's alpha > 0.80
Validity	
Convergent and discriminant validity	High correlations between scales measuring similar concepts Low correlations between scales measuring dissimilar
Responsiveness	Large (0>0.80) to moderate effect sizes (= 0.50)

Table 2 Characteristics of some MS-specific HRQoL instruments

	MSQOL-54	**FAMS**	**MSQLI**	**MSIS-29**	**Leeds MSQoL**
Original publication	Vickery et al., 1995	Cella et al., 1996	Ritvo et al 1997	Hobart et al., 2001	Ford et al., 2001
No of items	54	59 (of which 44 are scored)	137 (80 with abbreviated scales)	29	8
Dimensions	SF-36 and items in the following areas: health distress, sexual function, satisfaction with sexual function, overall quality of life, cognitive function, energy, pain and social function	mobility, symptoms, emotional well-being (depression), general contentment, thinking/fatigue, family/social well-being	9 scales including SF-36, Modified Fatigue Impact Scale, MOS Pain Effects, Sexual Satisfaction Scale, Bladder Control Scale, Bowel Control Scale, Impact of Visual Impairment, Perceived Deficits Questionnaire, Mental Health Inventory, MOS Social Support Scale	physical and psychological	unidimensional measure of well-being
Reliability	Internal consistency (0.75-96); Test-retest (0.66-96) [21]	Internal consistency (0.82-96); Test-retest (0.85-91) [22]	Internal consistency (0.75-0.97) [26]	Internal consistency (0.91-0.96); Test-retest (0.87-0.94) [27]	Internal consistency (0.79); Test-retest (0.85) [25]
Validity	Correlations among subscales conformed to predicted relationships among their underlying constructs Significant associates between MSQOL-54 scales and degree of multiple sclerosis symptom severity in the prior year, level of ambulation, employment limitations due to health, admission to hospital in the previous year, and depressive symptoms.[21]	Construct validity of the scales were supported by the predictable patterns of correlations among its subscales, and by relationships between its subscales and other measures (SF-36, Hospital Anxiety and Depression Scale, Multiscale Depression Inventory). FAMS subscales correlated predictably with self-assessed physical impairment (Eastern Cooperative Oncology Group Performance Status Rating), and for a subsample, mobility scores were correlated with EDSS and Scripps Neurologic Rating Scale. [22]	Construct validity supported of both the generic scale (SF-36) and the symptom-specific measures were supported by intercorrelations among scales. As expected, most symptom-specific measures correlated relatively weakly as expected with objective measures of impairment, as there was little correspondence between most symptom-specific scales and the impairment measures. One exception to this were the Impact of Visual Impairment Scale, which correlated moderately with the expected objective measures (ie visual acuity, Visual and Brainstem FSS, and EDSS) [26]	Direction, magnitude and pattern of correlations are consistent with predictions. For example, MSIS-29 physical scale correlated most highly with the FAMS mobility scale, the SF-36 physical functioning scale and the BI, and least with the EQ-5D anxiety/depression dimension, the SF-36 role emotional scale and the FAMS family/social well-being scale. Mean MSIS-29 scores for people who were retired due to MS were significantly higher than for those who were still employed [27]	Correlations with well-being (0.83) than to physical function (–0.39) LMSQoL able to detect significant difference between the ‘early relapsing remitting’ plus ‘benign’ groups and the ‘progressive’ groups [25]