The use of Multiple Sclerosis condition-specific measures to inform health policy decision-making: Mapping from the MSWS-12 to the EQ-5D

Running title: Mapping from the MSWS-12 to the EQ-5D

A Hawton, Health Economics Group, PenCLAHRC, Peninsula College of Medicine and Dentistry, University of Exeter, Veysey Building, Salmon Pool Lane, Exeter, Devon, UK EX2 4SG. 01392 722284; annie.hawton@pms.ac.uk

C Green, Health Economics Group, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK

CJ Telford, Health Economics Group, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK

DE Wright, Centre for Health & Environmental Statistics, University of Plymouth, Plymouth, UK

JP Zajicek, Clinical Neurology Research Group, Peninsula College of Medicine and Dentistry, University of Plymouth, Plymouth, UK

Multiple sclerosis; Cost effectiveness; Quality of life; Outcomes research; Walking; Decision Making

Abstract

Background: Walking impairment has a major influence on the quality-of-life of people with multiple sclerosis (MS). The Multiple Sclerosis Walking Scale (MSWS-12) assesses the impact of MS on walking ability from the patient's perspective, but in its current form, is not amenable for use in many policy decision-making settings.

Objectives: Statistical 'mapping' methods were used to convert MSWS-12 scores to EQ-5D health state values.

Methods: The relationship between the measures was estimated using cohort data from people with MS in South West England. Regression analyses were conducted, estimation errors assessed, and predictive performance of the best models tested using longitudinal data.

Results: Model performance was in line with that of other mapping studies, with the best performing models being an ordinary least squares (OLS) model using MSWS-12 item scores, and an OLS model using the total MSWS-12 score and its squared term.

Conclusions: A process has been described whereby data from a patient-reported outcome measure (MSWS-12) can be converted to (EQ-5D) health state values. These values may be used to consider the health-related quality-of-life of people with MS, to estimate quality-adjusted life-years for use in effectiveness and cost-effectiveness analyses, and to inform health policy decisions.

Introduction

Walking impairment is the most visible sign of multiple sclerosis (MS) [1] [2], with walking performance contributing greatly to the impact of MS on the health-related quality-of-life of people with the condition [3] [4]. Clinical measures commonly used to assess the progression of MS, such as the Expanded Disability Status Scale (EDSS) [5] and timed walk tests [6], do not directly inform on the impact of walking impairment on individuals' daily living [7]. Against this backdrop, and the wider move to the use of patient-reported outcome measures (PROMs) [8] [9], the Multiple Sclerosis Walking Scale (MSWS-12) has been developed to assess the impact of MS on walking ability from the patient's perspective [7]. The MSWS-12 is increasingly used in clinical trials, specifically where interventions are targeted at alleviating walking impairment (e.g. fampridine [10]; nerispirdine [11]; core stability training [12]; dynamic foot orthoses [13]). However, in its current form, it does not provide a preference-based measure of health status. This means it is not able to be used in many policy settings where decisions are informed by information on the cost-effectiveness of treatments.

Most measures of health status are non-preference-based [14]. They describe or assess the health state that an individual is in, but assign no value (or preference) to the state. These measures may give a summative or detailed picture of an individual's health, but give no indication of the relative value that would be given to this health state compared to other possible health states. However, decision-makers assessing value for money using comparative effectiveness or cost-effectiveness analyses [15] [16], generally require summary preference-based measures of health status. Such

measures use preference data, elicited from appropriate populations, to assign values to health state descriptions.

Preference-based measures have two components: i) a means of describing health status and; ii) a mechanism for assigning health state values to each of the possible health states [14]. The health state values, estimated from, most frequently, preferences of the general population, can be derived by a variety of methods, and typically give values on a scale where 1 is equivalent to full health and 0 is equivalent to death. Data from preference-based measures are more amenable for use in decision-making regarding the cost-effectiveness of interventions, as preferences for the health states, or the outcomes associated with interventions, can be compared across conditions [17]. In addition, preference-based measures are used to estimate quality-adjusted life-years (QALYs), increasingly the outcome of choice in policy settings involving the use of economic evaluations [18] [19] [20].

QALYs combine quantity and quality of life in a single measure of health outcome, by adjusting life years survived using a quality of life weight, with the weights usually being health state values derived from preference data [21]. For example, a year in full health would equate to 1 QALY and two years in ‘half health’ (0.5 health state value) would also equate to 1 QALY.

Preference-based measures, one of the most commonly used being the EQ-5D index [22], are now widely used in health policy settings. In contrast, the MSWS-12 is a non-

preference-based measure, meaning QALYs cannot be calculated from it and it is not directly applicable for use in cost-effectiveness analyses.

Historically, clinical trials have typically favoured condition-specific measures (such as the MSWS-12) as their outcome of choice, arguing that they are more likely to be sensitive to change than generic health status measures, such as the EQ-5D [14]. This has resulted in studies that provide data that can be used for assessing effectiveness, but not for assessing cost-effectiveness. This is the case with the MSWS-12.

There is an argument for the use of both a PROM and a preference-based measure in trials. The PROM will be particularly relevant to people with MS, more specific to the condition and, potentially, more sensitive to change, whilst the preference-based measure can be used for estimating QALYs and in cost-effectiveness analyses. However, in practice, a preference-based measure, such as the EQ-5D is often not included, resulting in limitations to the economic evaluation that can be conducted. This paper used statistical techniques to ‘map’ from the MSWS-12 to preference-based scores on the EQ-5D. The aim was to derive conversion algorithms that could be used in practice to convert MSWS-12 scores to EQ-5D scores for use in cost-effectiveness analyses.

Methods

The data

Data from the UK South West Impact of Multiple Sclerosis (SWIMS) project were used for analysis. SWIMS is a longitudinal, prospective, cohort study of people with MS in Devon and Cornwall (South West England), with individuals followed-up six-monthly [23]. Data are collected on demographics and clinical features, and across a range of patient-reported outcomes, including the MSWS-12 and the EQ-5D. SWIMS commenced recruitment in August 2004, and all participants who had completed baseline questionnaires including complete MSWS-12v1, EQ-5D and demographic (age and gender) data at February 2010 were included in this analysis.

The SWIMS study was approved in the UK by the Cornwall and Plymouth and South Devon Research Ethics Committees, and written informed consent obtained from all participants.

Measures

MSWS-12 [7] – The MSWS-12 is a patient-based measure assessing the extent to which an individual's MS impacts on their walking ability. It is a condition-specific measure, developed from people's experiences of MS [7], and has been validated using comprehensive psychometric techniques [24] [25]. The scale comprises 12 items rated on a five point scale (1, 'not limited' to 5, 'extremely'). Total scores are calculated and can range from 12 to 60. These scores are transformed to a scale of 0 to 100 to aid interpretation. Higher scores reflect greater impact on walking ability.

EuroQol EQ-5D [22] – The EQ-5D is a generic health status measure, comprising five sub-scales (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), with each sub-scale having three response levels (1, 'no problems'; 2, 'moderate problems'; 3, 'severe problems'). This classification of health status results in 243 possible health state descriptions.

Participant responses to the EQ-5D can be converted to the EQ-5D derived single index, a generic preference-based measure, using preference-weights for the health states. For example, in the UK, a tariff is commonly used [26] which has been derived from the preferences of a general population sample for each of the 243 possible health states. This gives values for each of the EQ-5D health states on an index ranging from 1.00 for the best health state to -0.594 for the worst health state. The EQ-5D is frequently used in clinical studies and cost-effectiveness analyses, and is currently the measure preferred by the UK National Institute of Health and Clinical Excellence in its health technology appraisals process [27].

Mapping

Over the last 10 years, statistical procedures have been developed for mapping (or 'cross-walking') from condition-specific measures, such as the MSWS-12, to generic preference-based measures, such as the EQ-5D index. Mapping is possible where data exist on the two measures from the same sample, and involves estimating the relationship between the measures using statistical association. Regression analyses are used to derive algorithms that can be used to convert non-preference-based scores

to preference-based scores (QALY weights). A wide range of studies have been published reporting on mapping from, for example, cancer [28], osteoarthritis [29], oral health [30], and Crohn's disease [31] condition-specific instruments, to preference-based measures such as the EQ-5D. However, little has been done in the field of neurology, with only one paper mapping from the Parkinson's Disease Questionnaire (PDQ-8) to the EQ-5D [32], and no work identified mapping from MS instruments to health state values.

Statistical analyses

Statistical conventions in the mapping literature [33] were followed to examine the relationship between the MSWS-12 and the EQ-5D index.

Baseline data from SWIMS were used as the estimation sample, to develop the most appropriate statistical models and to test within-sample performance. The accuracy of the best performing models in estimating EQ-5D scores was then assessed using longitudinal data from SWIMS.

Five regression models were initially estimated using SWIMS baseline data.

The EQ-5D index was regressed on the:

Total score for the MSWS-12 (Model A);

Total score for the MSWS-12 and the total score squared[1] (Model B);

Total score for the MSWS-12 and participant age and gender (Model C);

[1]The squared terms were added to pick up non-linearities in the relationship between MSWS-12 and health state value scores.

MSWS-12 item scores[2] (Model D);

MSWS-12 item scores and participant age and gender (Model E).

As described above, ratings on each of the MSWS-12 items are on a 5-point response scale. Mapping studies have mostly assumed that such response scales represent interval scale data, and have used linear regression analyses ([32], [29], [30], [34]. Whilst this approach is also adopted here, in recognition that the response options are ordered categories, the item responses were also re-categorised to a simple binary response scale, with categorical data as 0 ‘not at all’ and 1 ‘a little’/’moderately’/’quite a bit’/’extremely’ (in line with others e.g. [32], [30]), and the following models were also run:

Dichotomised MSWS-12 item scores (Model F)

Dichotomised MSWS-12 item scores and participant age and gender (Model G).

Ordinary least squares (OLS) regression models were used for estimation, together with Tobit [35] and Censored Least Adjusted Deviation (CLAD) [36]. These latter approaches were applied to address common concerns over OLS methods, and to test their value in further reducing estimation errors. Tobit models use an upper censoring limit of 1, as is the case with the EQ-5D, but are sensitive to violations of hetereoscedasticity or non-normality; in such instances, CLAD may be more appropriate as it is robust to these violations.

[2]Backward stepwise approach: The least statistically significant of the12 items were removed one by one until t he estimation errors increased.

When mapping, the principal aim is not to produce a model that explains the most variance in the data (adjusted R^2), but to derive an algorithm that as accurately as possible estimates health state values at a group level [37]. Therefore, for each of the regression models (21 in total), estimation errors [33], in the form of mean absolute error (MAE) and root mean square error (RMSE) were assessed.[3]

There are currently no guidelines as to when estimation errors are and are not acceptable [38], but a systematic review of mapping studies [33] has reported MAEs from 0.0011 to 0.19, and RMSEs from 0.084 to 0.2. (Adjusted R^2 values of 0.17 to 0.51 have been reported). Also assessed were the proportions of estimates that fell within 0.10 and 0.25 of the actual EQ-5D value.

The performance of the statistical models with the lowest estimation errors was then assessed with the SWIMS longitudinal data. Estimation errors were explored with the follow-up data and assessed according to the severity of the EQ-5D health state, and the actual EQ-5D health state values were compared with estimated values.

Data analysis was conducted in STATA 10.

Results

Participants

[3]MAE is the mean of the absolute estimation errors across individuals (the estimation error is the difference between the actual EQ-5D score for a particular individual and their estimated EQ-5D score based on the mapping model). RMSE is the positive square root of the mean squared estimation error.

560 SWIMS participants provided MS diagnosis, age, gender and MSWS-12v1 and EQ-5D data at baseline (Table 1), with the demographics of the sample being similar to published data for the UK [23]. The correlation between MSWS-12 total scores and EQ-5D health state values was r=-0.581 (p<0.001).

Table 1: Demographic and clinical characteristics of the SWIMS sample at baseline

Characteristic	*N = 560*
MS diagnosis, n (%):	
Relapsing-remitting	227 (40.5)
Primary progressive	104 (18.6)
Secondary progressive	103 (18.4)
Benign	17 (3.1)
Not known	109 (19.5)
Gender, n (%):	
Male	148 (26.4)
Female	412 (73.6)
Age (years), mean (sd)	50.2 (11.0)
(range)	(18 to 79)
MSWS-12 total score, mean (sd)	60.1 (32.4)
(range)	(0 to 100)
EQ-5D index score, mean (sd)	0.614 (0.248)
(range)	(-0.239 to 1)

Predictive performance (estimation sample)

In total 21 models were considered, with Models A to G each estimated using the three regression types.

Multi-collinearity of the MSWS-12 items was assessed using the 'collin' command in STATA. As a result, item 11 ('Affected how smoothly you walk') was removed from the items analyses. Models A through E showed evidence of hetereoscedasticity according to White's test, so the 'robust' command was used for the OLS and Tobit analyses (CLAD is known to be resistant to violations of hetereoscedasticity).

The OLS and CLAD approaches resulted in smaller estimation errors than the Tobit models, but the CLAD method of estimation substantially over-estimated EQ-5D scores for people in poorer health states. As such, only the OLS models were considered further.

The best performing model, Model D (MAE 0.148, RMSE 0.198), used MSWS-12 item scores. However, a practical concern is that a mapping algorithm based on item scores cannot be used when only summary data is available, so the best performing model that used aggregate data i.e. total scores (Model B) (MAE 0.150, RMSE 0.20) was also tested with the longitudinal data. In addition, the best performing model using dichotomised item scores (Model G) (MAE 0.150, RMSE 0.21) was tested with the follow-up data, as this model did not assume interval properties of the item responses. Details of these models are given in Table 2. The estimation errors were very similar across the three models (based on MAE, RMSE, percentage of estimates within 0.1 and

0.25 of the true EQ-5D value, and adjusted R^2), and were judged reasonable and in keeping with those found in other mapping studies.

Table 2: Predictive performance of best performing (OLS) mapping models

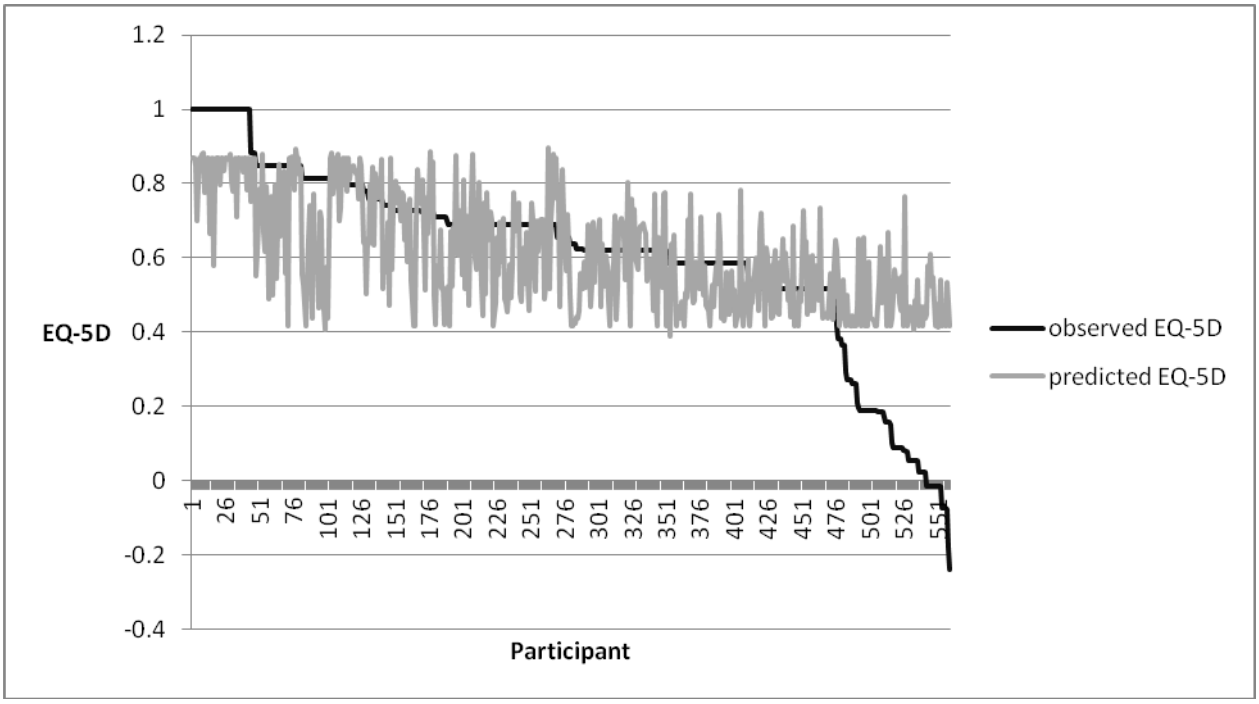
	Model B – MSWS-12 total & total score squared	Model D – MSWS-12 item scores†	Model G – Dichotomised items & demographics†
Estimation sample (n=560)			
Regression coefficients:			
MSWS-12 total score	-0.0047809***	-	
MSWS-12 total score squared	0.00000325	-	
MSWS-12 item 1	-	-0.0004282	-0.0491527
MSWS-12 item 2	-	-0.0029117	0.0739361
MSWS-12 item 3	-	-0.0213846*	-0.087689*
MSWS-12 item 4	-	-0.0410001***	-0.0707119
MSWS-12 item 5	-	0.0086472	-0.0226769
MSWS-12 item 6	-	-0.0340533**	-0.0237278
MSWS-12 item 7	-	-0.0154952	-0.0378606
MSWS-12 item 8	-	-0.0180406	-0.0769457*
MSWS-12 item 9	-	0.0004532	-0.0658229
MSWS-12 item 10	-	0.0148398	-0.0281462
MSWS-12 item 11	-	-	-
MSWS-12 item 12	-	-0.004274	0.0073948
Age	-	-	-0.0007418
Gender	-	-	0.005747
Constant	0.8863602	0.9843433	0.9418807
Predictive performance:			
Adjusted R^2	0.338	0.361	0.291
MAE (95% CI)	0.150 (0.139, 0.161)	0.148 (0.138, 0.159)	0.150 (0.138, 0.162)
RMSE	0.201	0.198	0.206
Estimates within $\pm$0.1 of true value (%)	44.82	46.79	50.54
Estimates within $\pm$0.25 of true value (%)	81.07	84.11	80.71
Longitudinal data (n responses=317)			
Predictive performance:			
MAE (95% CI)	0.163 (0.146, 0.181)	0.165 (0.148, 0.186)	0.172 (0.153, 0.191)
RMSE	0.227	0.229	0.241

Estimates within ±0.10 of true value (%)	47.95	47.32	50.47
Estimates within ±0.25 of true value (%)	77.60	78.55	76.03

*p<0.05, **p<0.01, ***p<0.001. †Removing the least statistically significant items did not improve the MAEs or RMSEs, so all items were retained (except item 11 due to multi-collinearity).

Figure 1 shows actual and estimated (from OLS items model D) EQ-5D scores for study participants, ranked by increasing EQ-5D score. This illustrates the problem of accurately estimating EQ-5D scores at, particularly the poor health end of the health severity spectrum. EQ-5D scores were over-estimated for the 14.8% of the sample who had scores less than 0.390. Scores were under-estimated for the 7.9% who scored greater than 0.895. (Over-estimation of EQ-5D scores for those in poor health was illustrated by the OLS dichotomised items Model G. This had a slightly higher percentage of EQ-5D estimates within 0.1 of the true value than Models B and D, but a slightly lower percentage of EQ-5D estimates within 0.25 of the true value than the other two models, as an apparent result of estimating marginally less accurately for the poorer health states).

Figure 1: Pattern of actual and estimated EQ-5D scores in the estimation sample for the OLS items model (D)



Predictive performance (longitudinal data)

Across six, 12 and 18 month follow-ups, 317 paired responses on the MSWS-12 and the EQ-5D were available. This data was used to assess estimation errors, according to the algorithms from the best performing models (given in Table 3).

Table 3: Algorithms of best performing models from estimation sample

Model	*Estimated EQ-5D score*
OLS items model (D)	= 0.9843433 - 0.0004282*MSWS-12 item 1 - 0.0029117*item 2 - 0.0213846*item 3 - 0.0410001*item 4 + 0.0086472*item 5 - 0.0340533*item 6 - 0.0154952*item 7 - 0.0180406*item 8 + 0.0004532* item 9 + 0.0148398*item 10 - 0.004274*item 12

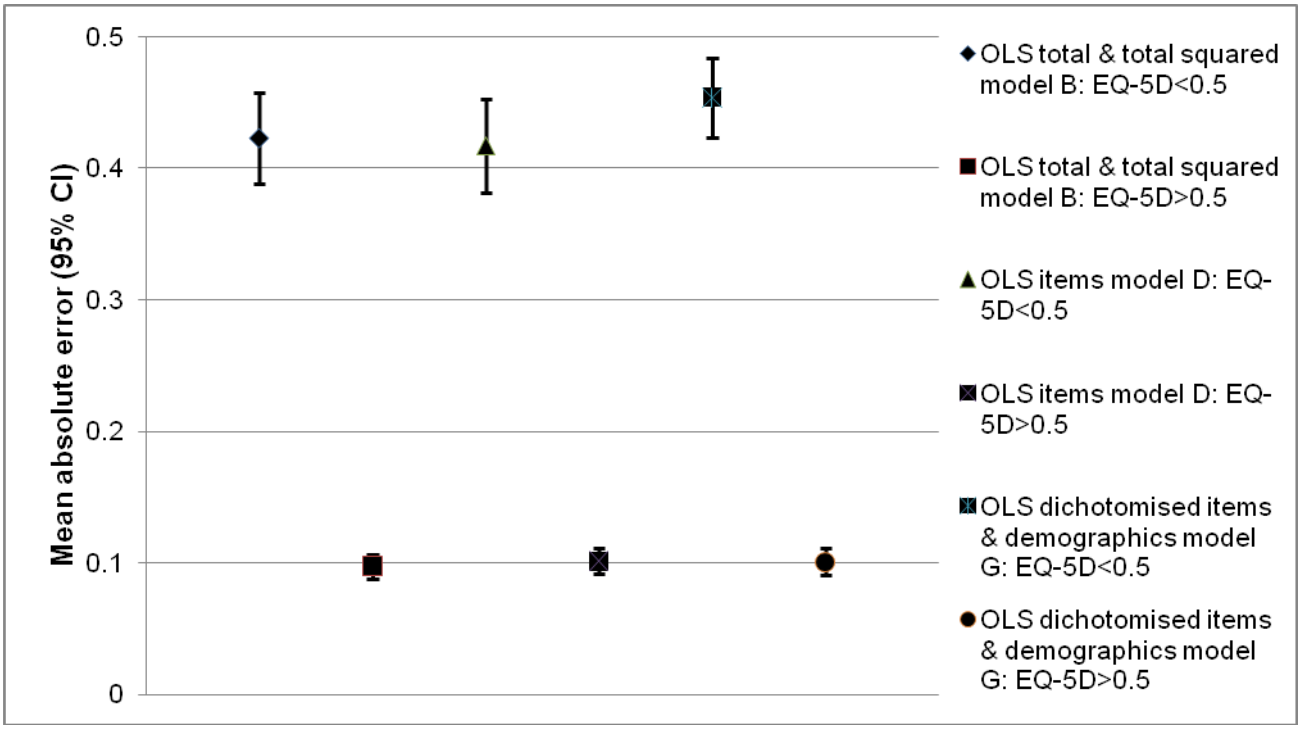
OLS total score & total score squared model (B)	= 0.8863602 - 0.0047809*MSWS-12 total score + 0.00000325*MSWS-12 total score*MSWS-12 total score
OLS dichotomised items & demographics model (G)[‡]	= 0.9418807 - 0.0491527*MSWS-12 dichotomised item 1 + 0.0739361*dichotomised (d)item 2 - 0.087689*(d)item 3 - 0.0707119*(d)item 4 - 0.0226769*(d)item 5 - 0.0237278*(d)item 6 - 0.0378606*(d)item 7 - 0.0769457*(d)item 8 - 0.0658229*(d)item 9 - 0.0281462*(d)item 10 + 0.0073948*(d)item 12 - 0.0007418*age at baseline + 0.005747*gender

[‡]The dichotomised items were scored 'not at all' = 0 and 'a little'/'moderately'/'quite a bit'/'extremely' = 1. Gender was coded 'male' = 0, 'female' = 1

The actual mean (sd) EQ-5D score in the follow-up sample was 0.571 (0.278), and the estimated scores from the mapping models B, D and G were 0.611 (0.140), 0.608 (0.147) and 0.607 (0.133), respectively. Estimation errors were slightly increased as compared to errors in the estimation sample (as would be expected) (Table 2), but were still in line with those found in other studies.

The analysis of estimation errors by health state severity showed a clear pattern. Figure 2 shows estimation errors (MAEs) for 'poor' (EQ-5D$\leq$0.5) and 'good' (EQ-5D>0.5) health states by model, and indicates that all the models performed less well at estimating EQ-5D scores for those in poorer health states.

Figure 2: Mean absolute errors by EQ-5D health state severity in the longitudinal sample



Additional modelling (2-part and 3-part models) and further data exploration

As a result of the reduced performance when estimating EQ-5D scores for the poorer health states, a range of additional models were run to explore whether optimal thresholds, or cut-offs, could be found, above and below which more accurate mapping algorithms could be determined.

The estimation sample was divided into two groups by MSWS-12 total score, and each of the model specifications (A to G) run on both groups. Multiple splits on the MSWS-12 total score (e.g. 90, 85, 80, 75 etc) were tested. A theoretically driven cut-off point on the MSWS-12 total score was also tested, following the approach of Versteegh [38]. The EQ-5D has a bi-modal distribution, with a 'gap' in the distribution at approximately 0.5. The median MSWS-12 total score of those with an EQ-5D score closest to 0.5 was

determined, and used as the cut-point for testing all model specifications. In addition, data were divided into two groups by EQ-5D score, and all model specifications tested on both groups. This was conducted for multiple cut-offs on the EQ-5D. Three-part models were also considered by dividing the data into three groups according to the EQ-5D scores for which the best performing model did and did not estimate (i.e. EQ-5D scores: <0.390; ≥0.390 and ≤0.895; >0.895).

The findings were the same for all of the two-part and three-part models; the errors could not be improved when estimating EQ-5D scores for those with poorer walking ability or those in poorer health states.

The MSWS-12 and EQ-5D profiles of those for whom the best performing model did and did not over-estimate EQ-5D scores (EQ-5D<0.390, EQ-5D≥0.390, respectively) were explored, by comparing the MSWS-12 total scores, item scores and EQ-5D dimension scores of these two groups. There were no apparent differences in MSWS-12 total or item scores, and the model estimated EQ-5D scores for individuals across the full spectrum of Walking Scale scores. When comparing the EQ-5D dimension profiles of the two groups, it was notable that a higher proportion of those for whom the algorithm over-estimated were unable to perform usual activities and had extreme pain/discomfort as compared to the other EQ-5D dimensions.

Discussion

Whilst the availability of new treatments for MS is welcome, the associated cost of new drugs creates tension between funding agencies, clinicians, patients and pharmaceutical companies over decision-making regarding cost-effectiveness [39], with the need to demonstrate 'value for money' of these treatments [16, 18].

To date, the reporting of health state values (such as those given by the EQ-5D) for people with MS has been linked only to clinical measures, particularly the EDSS [40]. Psychometrically-sophisticated patient-reported outcome measures (PROMs), which assess the impact of conditions on people's daily lives, are increasingly being developed in the field of neurology [9] [41], and are being used in clinical trials to assess the *effectiveness* of treatments [15]. However, there is a paucity of information regarding the relationships between PROMs and health state values [40] and, in their current form, condition-specific PROMs, such as the MSWS-12, are not amenable for use in policy decisions regarding *cost-effectiveness*, as they are not preference-based.

The work described here demonstrates a way in which PROMs can be linked to health state values which, in turn, may be used in considering the effectiveness and cost-effectiveness of new and existing treatments for MS. Data collected using a MS-specific measure (the MSWS-12) can be 'converted' to (EQ-5D) health state values, allowing estimation of QALYs and consideration of outcomes more broadly in a policy setting.

Limitations and strengths

The approach of using a statistically-derived algorithm to estimate health state values from a condition-specific measure is recognised as having limitations [33]. Mapping can only ever encompass the overlap between the descriptive systems of the two measures (e.g. the MSWS-12 and the EQ-5D). Information will be lost in the conversion process, leading to uncertainty about the preference for health that is actually included in health state value scores. Yet, approaches to considering uncertainty in mapping have focused entirely on empirical uncertainty (how good the model is at predicting scores), rather than theoretical uncertainty (what degree of information loss occurs in the mapping process) [42]. As Parker and colleagues [43] highlight, theoretical guidance to steer such analyses is limited, and this remains an area of health economics in which empiricism largely rules. This is particularly relevant when a *domain*-condition-specific measure, such as the MSWS-12, which is also only applicable to people who have some walking ability, is mapped to health state values, as there is likely to be less information overlap than between, for example, a generic measure of health status, and health state values. In addition, the sound psychometric properties of the MSWS-12 are weakened when mapped to EQ-5D scores. This said, theoretical uncertainty and psychometric sophistication may have limited relevance in the practical context of decision making where mapping algorithms can provide useful information. If health state values can be estimated from condition-specific measures (as is the case with the MSWS-12 and the EQ-5D), the theoretical uncertainty of what is lost in the mapping algorithm, whilst important to acknowledge, may become less relevant.

In addition, where preference-based data are not available to inform decisions, current alternatives to the use of statistical mapping methods are time consuming or impractical, and also open to criticism. For example, converting scores from PROMs to preference-based scores has also been conducted using the judgment of 'experts'. However, this approach has been criticised for its arbitrariness, and because no attempt is made to estimate the uncertainty around the conversion [33]. A further alternative to mapping is to elicit health state values from general population samples for each of the MSWS-12 health states. This would be a resource intensive process, with some methodological concerns, meaning that the expediency of mapping is increasingly used.

Implications

Until preferred alternatives are developed to enable the use of MSWS-12 and, more generally, MS-specific outcome measure, data in cost-effectiveness analyses and decision-making, we have shown that OLS regression statistical mapping, can provide algorithms that can be applied to MSWS-12 data for estimating EQ-5D scores. Algorithms have been generated that can convert MSWS-12 total scores, MSWS-12 item scores, and MSWS-12 dichotomised scores to EQ-5D health state values, each of which estimate EQ-5D scores with similar degrees of accuracy and give estimation errors (and R^2 values) in keeping with those found in other mapping studies [33]. Although, we acknowledge that there is currently no clear criteria as to when mapping is and is not acceptable [38].

The SWIMS sample is heterogeneous, as evidenced by the breadth of ages, mix of diagnoses, and wide range of EQ-5D scores. The demographics and clinical characteristics of the sample are similar to published data for the UK [23], and this supports the generalisability of the mapping algorithms described here to other populations of people with MS.

The statistical algorithms were more prone to error when estimating EQ-5D scores over time and for the more severe health states, a common finding in other mapping studies [37]. Additional work exploring two-part models did not ameliorate this over-estimation (in line with what has been found in other studies e.g. [44]) and, the newer approach of using 3-part models also did not improve the fit of scores. However, in the context of economic evaluations, comparisons are primarily made across groups and *individual* estimates are less important than the effect of estimation errors at the *group* level [37]. In the data used here,14.8% of the sample had scores below 0.390 (the minimum score predicted by the best performing models), meaning these errors may have little impact on the comparison of group level data [37], especially in specific evaluation settings where the target group are in the mild-to-moderate spectrum of disease severity.

Future directions

Further work is required to explore how EQ-5D scores estimated from MSWS-12 scores (using the algorithms provided) function as compared to actual EQ-5D scores, when calculating QALYs and the cost-effectiveness of interventions for people with MS. This research will investigate whether using estimated, rather than actual scores, is likely to

make a difference in practical policy contexts regarding decisions as to whether to fund treatments, or not.

Conclusions

A process has been described whereby data from a condition-specific outcome measure, particularly relevant to people with MS, can be converted to health state values. These health state value estimates have the potential for use in cost-effectiveness analyses of treatments for people with MS and can inform the ongoing health policy debate regarding such interventions [45], and their value for money [16].

Acknowledgements

The South West Impact of Multiple Sclerosis (SWIMS) project has been supported through funding from the Multiple Sclerosis Society of Great Britain and Northern Ireland, the Peninsula Medical School Foundation and the UK NIHR Comprehensive Clinical Research Network. This article presents independent research funded by the National Institute for Health Research (NIHR). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

1. Heesen C, Bohm J, Reich C, Kasper J, Goebel M, Gold S. Patient perception of bodily functions in multiple sclerosis: gait and visual function are the most valuable. *Multiple Sclerosis*. 2008; 14: 988-91.
2. Sosnoff J, Goldman MD, Motl RW. Real-life walking impairment in multiple sclerosis: preliminary comparison of four methods for processing accelerometry data. *Multiple Sclerosis*. 2010; 16: 868-77.
3. Royal College of Physicians MT. National audit of services for people with multiple sclerosis. London: RCP, 2008.
4. Finlayson M, Van Denend T. Experiencing the loss of mobility: perspectives of older adults with MS. *Disability & Rehabilitation*. 2003; 25: 1168-80.
5. Kurtzke J. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983; 33: 1444-52.
6. Fischer J, Jak A, Kniker J, Rudick R. Multiple Sclerosis Functional Composite (MSFC): administration and scoring manual. New York: National Multiple Sclerosis Society, 2001.
7. Hobart J, Riazi A, Lamping D, Fitzpatrick R, Thompson A. Measuring the impact of MS on walking ability: the 12-item MS Walking Scale (MSWS-12). *Neurology*. 2003; 60: 31-6.
8. US Department of Health and Human Services Food and Drug Administration. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Silver Spring, USA, 2009.
9. Miller D, Rudick R. Patient-centred outcomes: Translating clinical efficacy into benefits on health-related quality of life. *Neurology*. 2010; 74: S24-35.
10. Goodman A, Brown T, Krupp L, Schapiro R, Schwid S, Cohen R, et al. Sustained-release oral fampridine in multiple sclerosis: a randomised, double-blind, controlled trial. *Lancet*. 2009; 373: 732-38.
11. ClinicalTrials.gov. Efficacy, Safety and Tolerability of Nerispirdine in Patients with Multiple Sclerosis. 2011.
12. Freeman J, Gear M, Pauli A, Cowan P, Finnigan C, Hunter H, et al. The effect of core stability training on balance and mobility in ambulant individuals with multiple sclerosis: a multi-centre series of single case studies. *Multiple Sclerosis*. 2010; 16: 1377-84.
13. Ramdharry G, Marsden J, Day B, Thompson A. De-stablizing and training effects of foot orthoses in multiple sclerosis. *Multiple Sclerosis*. 2006; 12: 219-26.
14. Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press, 2007.
15. Dorsey E, Meltzer D. The economics of comparative effectiveness research. *Neurology*. 2010; 75: 492-3.
16. Rigel S, Swash M. Money and medicine: A problem that won't go away. *Neurology*. 2009; 72: 766-8.
17. Gold M, Siegel J, Russell L, Weinstein M. *Cost-Effectiveness in Health and Medicine*. Oxford: Oxford University Press, 1996.

18. Rawlins M, Culyer A. National Institute for Clinical Excellence and its value judgments. *British Medical Journal*. 2004; 329: 224-7.
19. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3). Barton, Australia: Australian Government. Department of Health and Ageing, 2008.
20. Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the economic evaluation of health technologies. 3rd edition. Ottawa, Canada, 2006.
21. Glick H, Doshi J, Sonnad S, Polsky D. *Economic Evaluation in Clinical Trials*. Oxford: Oxford University Press, 2007.
22. EuroQol Group. EQ-5D user guide. Rotterdam, Netherlands: The EuroQol Group, 1996.
23. Zajicek J, Ingram W, Vickery J, Creanor S, Wright D, Hobart J. Patient-orientated longitudinal study of multiple sclerosis in south west England (The South West Impact of Multiple Sclerosis Project, SWIMS) 1: protocol and baseline characteristics of cohort. *BMC Neurology*. 2010; 10: 88.
24. Motl R, Snook E. Confirmation and extension of the validity of the Multiple Sclerosis Walking Scale-12 (MSWS-12). *Journal of the Neurological Sciences*. 2008; 268: 69-73.
25. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technology Assessment*. 2009; 13: 1-216.
26. Dolan P. Modeling Valuations for EuroQol Health States. *Medical Care*. 1997; 35: 1095-108.
27. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence, 2008.
28. Crott R, Briggs A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *European Journal of Health Economics*. 2010; 11.
29. Xie F, Pullenyegum E, Li S-C, Hopkins R, Thumboo J, Lo N-N. Use of a Disease-Specific Instrument in Economic Evaluations: Mapping WOMAC onto the EQ-5D Utility Index. *Value in Health*. 2010; 8: 873–8.
30. Brennan D, Spencer A. Mapping oral health related quality of life to generic health state values. *BMC Health Services Research*. 2006; 6: 96.
31. Buxton M, Lacey L, Feagan B, Oliver R. Mapping from Disease-specific measures to utility: an analysis of the relationship between the Inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's disease and measures of utility. *Value in Health*. 2007; 10: 214-20.
32. Cheung Y, Tan L, Lau P, Au W, Luo N. Mapping the eight-item Parkinson's Disease Questionnaire (PDQ-8) to the EQ-5D utility index. *Quality of Life Research*. 2008; 17: 1173-81.
33. Brazier J, Yang Y, Tsuchiya A, Rowen D. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *European Journal of Health Economics*. 2010; 11: 215-25.
34. Brazier J, Kolotkin R, Crosby R, Williams G. Estimating a Preference-Based Single Index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) Instrument from the SF-6D. *Value in Health*. 2004; 7: 490-8.

35. Long J, Freese J. *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications, 1997.
36. Powell J. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*. 1984; 25: 303-25.
37. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health and Quality of Life Outcomes*. 2009; 7: 27.
38. Versteegh M, Rowen D, Brazier J, Stolk E. Mapping onto EQ-5D for patients in poor health. *Health and Quality of Life Outcomes*. 2010; 8: 141.
39. Phillips C, Humphreys I. Assessing cost-effectiveness in the management of multiple sclerosis. *ClinicoEconomics and Outcomes Research*. 2009; 1: 61-78.
40. Naci H, Fleurence R, Birt J, Duhig A. The impact of increasing neurological disability of mulitiple sclerosis on health utilities: a systematic review of the literature. *Journal of Medical Economics*. 2010; 13: 78-89.
41. Hobart J, Cano S, Zajicek J, Thompson A. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurology*. 2007; 6: 1094-105.
42. Round J. Capturing information loss in estimates of uncertainty that arise from mapping algorithms. London: Health Economists' Study Group, 2008.
43. Parker M, Haycox A, Graves J. Estimating the Relationship between Preference-Based Generic Utility Instruments and Disease-Specific Quality-of-Life Measures in Severe Chronic Constipation: Challenges in Practice. *Pharmacoeconomics*. 2011; 10.2165/11588360.
44. Dakin H, Petrou S, Haggard M, Benge S, Williamson I. Mapping analyses to estimate health utilities based on responses to the OM8-30 otitis media questionnaire. *Quality of Life Research*. 2010; 19: 65-80.
45. Prosser L, Kuntz K, Bar-Or A, Weinstein M. Patient and community preferences for treatments and health states in multiple sclerosis. *Multiple Sclerosis*. 2003; 9: 311-9.