Using the Fatigue Severity Scale to inform healthcare decision-making in multiple sclerosis: mapping to three quality-adjusted life-year measures (EQ-5D-3L, SF-6D, MSIS-8D)

Goodwin E PhD[1], Hawton A PhD[1,2], Green C PhD[1,2]

1 Health Economics Group, Institute of Health Research, University of Exeter, Exeter, UK

2 South West Collaboration for Leadership in Applied Health Research and Care (CLAHRC), University of Exeter Medical School, University of Exeter, Exeter, UK

Corresponding author: Annie Hawton, Health Economics Group, South Cloisters, St Luke's Campus, University of Exeter, Exeter UK EX1 2LU @HEG_Exeter

a.hawton@exeter.ac.uk

+44 1392722284

e.goodwin@exeter.ac.uk

c.green@exeter.ac.uk

Keywords: cost effectiveness, decision making, multiple sclerosis, outcomes research, quality of life, fatigue

Abstract

Background: Fatigue has a major influence on the quality of life of people with multiple sclerosis. The Fatigue Severity Scale is a frequently used patient-reported measure of fatigue impact, but does not generate the health state utility values required to inform cost-effectiveness analysis, limiting its applicability within decision-making contexts. The objective of this study was to use statistical mapping methods to convert Fatigue Severity Scale scores to health state utility values from three preference-based measures: the EQ-5D-3L, SF-6D and Multiple Sclerosis Impact Scale-8D.

Methods: The relationships between the measures were estimated through regression analysis using cohort data from 1056 people with multiple sclerosis in South West England. Estimation errors were assessed and predictive performance of the best models were tested in a separate sample (n=352).

Results: For the EQ-5D and the Multiple Sclerosis Impact Scale-8D, the best performing models used a censored least absolute deviation specification, with Fatigue Severity Scale total score, age and gender as predictors. For the SF-6D, the best performing model used an ordinary least squares specification, with Fatigue Severity Scale total score as the only predictor.

Conclusions: Here we present algorithms to convert Fatigue Severity Scales scores into health state utility values based on three preference-based measures. These values may be used to estimate quality adjusted life-years for use in cost-effectiveness analyses and to consider the health-related quality of life of people with multiple sclerosis, thereby informing health policy decisions.

Background

Over the last two decades, various disease-modifying and symptomatic treatments have been developed for people with MS. Meanwhile, increasing emphasis has been placed on achieving “value for money” within healthcare systems (1). Clinical trials of interventions that target particular symptoms frequently use symptom-specific outcome measures in order to maximise sensitivity and responsiveness to change. Fatigue is the most common symptom experienced by people with MS, and has a considerable impact on quality of life (2). The Fatigue Severity Scale (FSS) (3) is frequently used in clinical trials of interventions for fatigue in people with MS, including carnitine, amantadine, aspirin, modafinil and cognitive behavioural therapy (4) (5) (6) (7). Symptom-specific outcome measures, such as the FSS, provide a standardised means of describing “health states” that may be experienced by patients, but do not provide data in the format required by many decision-making bodies to assess cost-effectiveness (1).

The quality-adjusted life-year (QALY) is recommended for use as an outcome measure for cost-effectiveness analyses by several national decision-making bodies, eg the National Institute for Health and Care Excellence (NICE) (8) (9) (10). QALYs combine quantity and quality of life in a single measure, by adjusting the number of life-years lived according to the quality-of-life experienced during those years (1). In order to estimate QALYs, numerical values must be assigned to reflect the quality of life experienced when living in particular health states. These values are commonly obtained using preference-based measures (PBMs) of health-related quality of life (11).

However, many clinical trials do not include a PBM, limiting the ability to conduct economic evaluations. In such cases, statistical procedures may be used to “map” scores on non-preference based outcome measures, such as the FSS, to HSUVs derived from PBMs. “Mapping’ involves regression analysis, using a dataset containing responses to both measures from the same sample, to derive an algorithm that can be used to convert data from non-preference-based measures into HSUVs. Over recent years, the use of mapping has increased considerably (11). Previous studies have reported on mapping from MS-specific outcome measures including the Multiple Sclerosis Impact Scale and the Multiple Sclerosis Walking Scale-12 (12) (13) (14). However, no approach has been reported that uses fatigue measures to map to HSUVs in the context of MS.

Methods

This paper uses statistical techniques to map from the FSS (the “source measure”) to HSUVs derived from three preference-based measures: the EQ-5D, SF-6D and MSIS-8D (the “target measures”). The aim is to derive algorithms to convert FSS scores into HSUVs for use in assessing the cost-effectiveness of treatments for fatigue in people with MS. The statistical approach presented here is based on good practice methodology, and is consistent with the recommendations regarding mapping methods from NICE in the UK (15) and the international ISPOR Good Practices for Outcomes Research Task Force (16).

Measures

The Fatigue Severity Scale (FSS) has acceptable reliability, internal consistency, sensitivity and responsiveness for people with MS (3) (17) (18) (19) (20) (21). It comprises nine statements, describing the severity and impact of fatigue, with a scale of possible responses ranging from 1 (“strongly disagree”) to 7 (“strongly agree”). FSS total scores are usually reported as the mean score over the nine items; a higher score indicates greater severity.

The EuroQoL EQ-5D-3L has five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) with three response levels per dimension - no problems, some problems or extreme problems/confined to bed. HSUVs were derived from the preferences of a representative sample of the UK general population, using a variant of the time trade-off (TTO) technique, and range from -0.594 to 1.000 (22). The EQ-5D is widely used in economic evaluation, particularly in the UK, where NICE recommend it as the preferred measure of health outcomes for cost effectiveness analyses (8).

The Short-Form 6D (SF-6D) enables HSUVs to be estimated from a popular non-preference based measure of HRQoL, the Short-Form 36 (SF-36). It consists of six dimensions (physical functioning, role limitations, social functioning, pain, mental health, vitality) with between four and six response levels. Preferences were elicited from a representative sample of the UK general population using the standard gamble technique and values range from 0.301 to 1.000 (23). The SWIMS dataset includes responses to Version 1 of the SF-36 from earlier waves of data collection, before this was replaced by SF-36 Version 2, which was developed to address concerns about the structure and wording of some items (24). Given that the component items of the SF-6D classification system differ between the two versions, we only included responses to Version 2 of the SF-36 in this analysis, in order to ensure consistency.

The Multiple Sclerosis Impact Scale 8D (MSIS-8D) enables HSUVs to be estimated from responses to an MS-specific outcome measure, the Multiple Sclerosis Impact Scale (MSIS-29). It includes eight dimensions (physical function; social and leisure activities, mobility, daily activities, mental fatigue, emotional well-being, cognition, depression) with four response levels (25). HSUVs were derived from a TTO survey with a sample of the UK general population. Values range from 0.079 to 0.882. It was not assumed that the best health state described by the MSIS-8D classification system (ie “no problems” on all dimensions) was equivalent to perfect health, therefore the value of this health state was not constrained to 1 (26). The MSIS-8D was derived from Version 2 of the MSIS-29 (21), which has four response levels per item, rather than Version 1 of the MSIS-29, which has five response levels (27). Therefore, although earlier waves of SWIMS data collection used Version 1 of the MSIS-29, only responses to Version 2 were included in this analysis.

Dataset

The South West Impact of Multiple Sclerosis (SWIMS) project is a longitudinal cohort study of people with MS aged 18 or over, living in Devon and Cornwall. Respondents complete six-monthly questionnaires, including several patient-reported outcome measures alongside clinical and demographic characteristics. The study was approved in the UK by the Cornwall and Plymouth and South Devon Research Ethics Committees, and written informed consent is obtained from all participants.

This analysis used SWIMS data received between August 2004 and October 2012. Only data collected at baseline were included, as this is the only point at which the FSS, EQ-5D, SF-36 and MSIS-29 are completed simultaneously. A random sample of 75% of the baseline data were used as the estimation dataset (n=1056), with the remaining 25% constituting the validation dataset (n=352) (28) (11). As Table 1 shows, there were no significant differences ($p<0.05$) between the datasets in terms of mean FSS total scores, mean HSUVs, or recorded demographic or clinical characteristics. The mapping algorithms were derived using data from respondents who provided answers to all questions required to produce both a FSS total score and a HSUV from the target PBM: 1023 respondents for the EQ-5D, 607 for the SF-6D and 650 for the MSIS-8D (response numbers are lower for the SF-6D and the MSIS-8D as only version 2 of these questionnaires were included). All statistical analysis was undertaken in Stata 14.

Preliminary assessment of measures

Two key conditions must be met for mapping: there should be conceptual overlap between the source and target measures, and the target measure should demonstrate discriminative validity with respect to the severity of the condition captured by the source measure (11) (29). To assess conceptual overlap, the FSS items and the dimensions of the PBMs were allocated to a multi-dimensional conceptual framework, which was developed for this study in order to provide a structure for comparing the content of the measures. The measurement concept underpinning the three PBMs is health-related quality of life (HRQL) (22) (23) (25). Therefore, the conceptual framework was structured around the commonly agreed key dimensions of HRQL, which comprise physical and mental domains alongside a third domain relating to social and role function and participation (30) (31) (32). The framework was constructed based on a systematic literature review of qualitative research into the impact of fatigue on people with MS (details of this review are included as Supplementary Material A). Pearson correlation coefficients were assessed between the total FSS score and HSUVs from each of the PBMs, while Spearman correlation coefficients were assessed between FSS total scores and individual dimension scores for each PBM, and between HSUVs and individual FSS item scores. Assuming that these instruments measure distinct but related concepts, we expected to find relationships of moderate strength, ie correlation coefficients between 0.3 and 0.6 (33). To assess the discriminative validity of the PBMs, respondents were categorised into fatigue severity groups: “mild/ no fatigue” (FSS total ≤ 35), “moderate fatigue” (36 ≤ FSS total ≤ 52) and “severe fatigue” (FSS total ≥ 53). The definition of “mild/ no fatigue” was based on the published cut-off point for the FSS (17). The ability of the PBMs to differentiate between the three groups was investigated using ANOVA and standardised effect sizes. Effect sizes can be assessed as small (0.20–0.49), moderate (0.50–0.79) or large (0.80 or over) (34).

Development of mapping algorithms

Exploration of model specifications

The relationships between the source and target measures were examined using statistical conventions reported in the mapping literature (29) (35). The distribution of scores on each of the measures was explored by the production of histograms and, the relationship between each of the PBMS and the FSS total score was investigated by production of scatterplots. Five regression models were estimated for each PBM. HSUVs were regressed on the:

- Total FSS score for the FSS (Model A);
- Total FSS score for the FSS and total FSS score squared (Model B);
- Total FSS score, age and gender (Model C);
- FSS item scores (Model D);
- FSS item scores, age and gender (Model E).

The majority of mapping studies estimate algorithms using ordinary least squares (OLS) models (35). However, OLS models can predict values outside the possible range for a PBM, and can lack predictive accuracy for extreme HSUVs. To address this, Tobit models were also considered, specifying an upper limit of 1 (29). OLS and Tobit models rely on an assumption of no heteroscedasticity. Where this assumption was violated according to White's test for heteroscedasticity, the 'vce(robust)' option was used in conjunction with the 'regress' command for the OLS analyses, and Censored Least Adjusted Deviation (CLAD) estimation methods (36) were used instead of Tobit models, employing the 'clad' command with a specified upper limit of 1.

Predictive ability was assessed using the following estimation errors: mean absolute error (MAE), root mean squared error (RMSE) and the proportions of estimates that fell within 0.05, 0.10 and 0.25 of the observed HSUV. MAE was selected as the primary criterion for selection of the preferred models (11). However, if coefficients had unexpected signs these models were not selected. In instances where model MAEs were the same, the model with the best profile of estimates falling within 0.05, 0.10 and 0.25 of the observed HSUV was selected.

Two researchers decided independently which models to would take forward for validation. Where discrepancies arose, these were resolved through discussion until consensus was reached. Demographic variables may not be included in the datasets from which HSUVs are to be estimated. Therefore, where the best performing models included demographic variables, the best performing model without demographic variables was also selected.

<u>Validation and model selection</u>

Estimation errors were assessed according to the severity of the health state. The selected models were applied to the validation dataset and their performance was assessed using the criteria outlined above.

Results

Preliminary assessment of measures

The conceptual framework that was developed to assess conceptual overlap between the measures is illustrated in Figure 1. Most of the themes that had been identified in the original qualitative research studies fitted into the three domains of HRQoL that were defined *a priori*. There were two notable exceptions. Several of the themes described the experience of fatigue itself, rather than its effect on HRQoL. This experience was clearly of great importance to the people with MS who contributed to the original research, and underpinned the ways in which fatigue impacts upon HRQoL. Therefore, an additional domain was added:

"Descriptions of fatigue". In terms of the links between themes, a clear relationship emerged between "functioning and participation" and "psychological well-being". People with MS specifically identified negative effects on their psychological well-being that were caused by the impact of their fatigue on their functioning and participation. These stood alongside, but distinct from, the direct impact of fatigue on psychological well-being. Therefore, this became a domain in its own right.

In terms of conceptual overlap, the FSS and all PBMs cover the three primary domains of the conceptual framework (Physical, Mental and Participation Effects) (Table 2). Coverage of Participation Effects is strong across all four measures. The FSS, SF-6D and MSIS-8D capture a wide range of Physical Effects, whereas the EQ-5D includes only specific dimensions for pain/discomfort and mobility. In terms of Mental Effects, the FSS includes one item relating to motivation, while the PBMs describe other specific symptoms eg depression or anxiety. Only the MSIS-8D includes cognitive effects. The MSIS-8D and SF-6D include dimensions relating specifically to fatigue or vitality.

Significant ($p<0.0001$) moderate correlations were evident between the FSS total score and HSUVs derived from the EQ-5D ($r = -0.455$) and the MSIS-8D (-0.590). There was a large significant correlation ($p<0.0001$) between the FSS total score and HSUVs derived from the SF-6D (-0.647). The FSS total score was significantly correlated with all individual dimensions of the PBMs, and HSUVs derived from each of the PBMs were significantly correlated with all individual items of the FSS ($p<0.0001$). Most correlations were moderate, as anticipated, and all had the expected negative sign, ie higher FSS scores are related to lower HSUVs (Table 3).

28.4% of respondents with a valid FSS total score were in the "mild/ no fatigue" category, 36.6% were in the "moderate fatigue" category and 35.0% were in the "severe fatigue" category. All PBMs discriminated significantly between fatigue severity groups ($p<0.0001$).

The SF-6D performed particularly well, with large standardised effect sizes (≥0.80). Overall, standardised effect sizes were higher for the MSIS-8D than for the EQ-5D (Table 4).

As a result of the preliminary assessments, it was judged that conceptual overlap and discriminative validity were sufficient to proceed with the estimation of mapping models. Overall, the SF-6D and MSIS-8D provide a better fit with the FSS.

Results of mapping analysis

Exploration of model specifications

In order to allow for heteroscedasticity, skewness and kurtosis identified in the data, we fitted robust OLS models and used a CLAD rather than a Tobit specification. (The distribution of scores on each of the measures, and the relationships between scores on the PBMs and the FSS total score is shown in the Supplementary Material B and C). Thirty models were considered, with Models A to E estimated for each PBM, using both OLS and CLAD specifications.

There was little difference between the predictive ability of the models based on FSS total scores and individual FSS items. In all models, item FSS-08 had a significant coefficient with an unexpected sign, and a majority of the FSS items (ranging from five to seven of the nine items) were not significant predictors of HSUVs. Furthermore, data on individual FSS items may not be available in all potential applications of the mapping algorithms. Therefore selection was restricted to algorithms based on the FSS total score.

EQ-5D: CLAD C had the lowest MAE and the highest proportion of individuals with small prediction errors. We also selected CLAD A, as the model which did not include demographic variables with the lowest MAE.

SF-6D: OLS B and CLAD B had coefficients with unexpected signs and were, therefore, not selected. We selected CLAD C as it had the next lowest MAE, and OLS A and CLAD A, as they did not include demographic variables.

MSIS-8D: CLAD B and OLS B had the lowest MAEs, however these had unexpected signs for FSS total, and so were not selected. The model with the next lowest MAE and highest proportion of individuals with small predictions errors was CLAD C. As this model included demographic variables, we also selected the model with the next lowest MAE (0.117), CLAD A.

Details of the selected models are presented in Table 5. All model results are provided in Supplementary Material D.

Validation and model selection

The validation dataset was used to assess estimation errors for the selected models (Table 6). Table 7 shows MAEs for 'poor' and 'good' health states by model. The models predicting HSUVs for the EQ-5D and MSIS-8D had larger MAEs for poorer health states, indicating that these models performed less well at estimating EQ-5D scores for those in poorer health states. The opposite was true for the SF-6D models, although the difference in MAEs here was less marked. (Please see Supplementary Materials E and F).

Discussion

Here we describe and demonstrate a method for converting responses to the FSS, a frequently-used measure of fatigue severity, into HSUVs, which can be used to estimate QALYs for use in cost-effectiveness analyses, and hence to inform decision-making regarding the availability of treatments for MS-related fatigue. According to the Oxford Health Economics Research Centre's Mapping Database, last updated in April 2019 (37), no previous published studies have attempted mapping from the FSS. In addition, we have found no previous studies which have investigated correlations between the FSS and the SF-6D or the FSS and the MSIS-8D, and just two which have explored the relationship between the FSS and the EQ-5D (38) (39). Rosa et al. (39) correlated FSS total scores with participants' scores on the EQ-5D visual analogue scale, rather than with the EQ-5D HSUVs that are relevant for mapping, and Tremmas et al. (38) found no statistically significant correlation between the FSS and EQ-5D scores of people with lung cancer.

The ability of the models selected in the current study to predict SF-6D and MSIS-8D values is in keeping with results reported in other mapping studies (35). There are currently no guidelines regarding acceptable limits for estimation errors (13), but MAEs ranging from 0.0011 to 0.19 have been previously described (35). In the current study, the SF-6D MAEs of 0.078 and 0.077 and the MSIS-8D MAEs of 0.117 and 0.116, fall well within this range and, specifically in the context of MS, they are in keeping with the MAE of 0.058 reported by Hawton et al. (12) when the MSIS-29 was mapped to the SF-6D.

Results for the EQ-5D algorithms were less convincing. The prediction errors of 0.175 and 0.173 are towards the higher end of MAEs reported in previous mapping studies (35), and are also high in the context of MS mapping studies. Versteegh et al. (13) mapped from the version 1 of the MSIS-29 to the EQ-5D, with a resulting MAEs of 0.13 and 0.16, and Hawton and colleagues (12) mapped from version 2 of the same measures to the EQ-5D with a MAE of 0.147. In addition, when testing the external validity of the Versteegh et al. (13) algorithm, Ernstsson et al. (40) reported a MAE of 0.12.

Information is inevitably lost in the process of mapping, as the resulting algorithm will only reflect the areas of content that overlap between the starting and target measures. This information loss is accentuated when a domain-specific, condition-specific measure, such as the FSS, is mapped to a generic, multi-dimensional measure, such as the EQ-5D. Therefore, greater predictions errors might be anticipated when mapping from such a uni-dimensional scale as the FSS than when mapping from a multi-dimensional scale such as the MSIS-29 (41). However, this does not appear to hold in the MS mapping literature to date, with Hawton et al. (14) reporting a MAE of 0.148 when they mapped from the MS Walking Scale-12 (a mobility-specific, MS-specific measure) to the EQ-5D, and Sidovar et al. (42) described an error statistic of 0.109 when mapping to/from these same measures.

In the current study, the EQ-5D algorithms were particularly problematic for HSUVs below 0.65. They did not predict any values below 0.54 (assuming an age of 50 years and female gender for CLAD Model C), which is of particular concern for a measure with a minimum value of -0.594.

On the basis of the statistical assessments reported here, the qualitative assessments of conceptual validity, and setting our findings in the context of other mapping studies in MS and mapping studies more generally, we suggest the use of the following algorithms for mapping from the FSS to HSUVs.

SF-6D estimate = 0.897 - 0.006*FSS total score

MSIS-8D estimate = 1.084 - 0.008*FSS total score – 0.001*age – 0.0024*gender [0 male, 1 female] or age and gender are not available:

MSIS-8D estimate = 0.985 – 0.007*FSS total score

Based on these same assessments, we suggest the EQ-5D algorithms are far less likely to produce accurate or valid estimates of EQ-5D scores.

There are a number of potential limitations of this work. Firstly, the SWIMS data were collected prior to the development and use of the EQ-5D-5L and the mapping algorithms were based on the ‘older’ EQ-5D-3L. It may have been expected that the EQ-5D-5L would supersede the EQ-5D-3L as it was developed with five, rather than the original three, levels in an attempt to improve its responsiveness. However, the English HSUV set for the EQ-5D-5L is not in common use, and if using the EQ-5D-5L descriptive system, the current ‘position statement’ of NICE is to use a cross-walk algorithm to provide HSUVs from the EQ-5D-3L value set. Secondly, the SF-6D value set is based on the use of standard gamble to elicit preferences for health states. This may result in higher HSUVs (than the EQ-5D), as respondents tend to be risk adverse. Thirdly, we did not explore the performance of some of the ‘newer’ mapping model specifications, such as limited dependent variable mixture models or beta-based regression, which may have better accounted for the bi-modal nature of the EQ-5D data. There is some empirical evidence in support of these models, but the ISPOR Task Force report (16) does not advocate any specific regression approach for mapping, recognising that the performance of different methods will vary dependent on a number of factors including the nature of the starting/target measures, the disease, and the patient population. The report suggests it is wise to use a model type for which there is existing evidence of good performance. In the context of MS, mapping algorithms which have used the same regression approaches that we have used here have been reported with MAEs of 0.058 (12), 0.13 and 0.16 (13), 0.147 (12), 0.12 (40), 0.148 (14) and 0.109 (42). Brazier et al.’s (35) systematic review of mapping studies reported MAEs of 0.0011 to 0.19. Therefore, the regression approaches in the current paper have a track record of use and acceptability in the context of MS. The MAEs reported here for the SF-6D and MSIS-8D are in keeping with those reported in these other mapping studies. The poor performance of the EQ-5D algorithms is likely to be a function of the limited conceptual overlap between the EQ-5D and the FSS. The limited shared conceptual content of these measures will not be altered by using a different form of regression analysis. Thirdly, algorithms to predict HSUVs from individual FSS items, rather than the total score, were not generated by this study. This was, in part, due to an anomaly affecting item FSS-08 (Fatigue is among the most disabling of my symptoms). While the item correlated negatively (as expected) with HSUVs when considered in isolation, it had a positive coefficient when included as an independent variable in regression analysis. Further research would be required to understand the mechanisms behind this; in the meantime, it is not possible to determine whether this item is suitable for inclusion in a mapping algorithm.

A particular strength of this study is the nature of the SWIMS dataset. It has provided comprehensive data on which to base the estimation and validation of these mapping algorithms. Importantly, the cohort is comparable with other UK-based samples of people with MS in terms of age, gender, relapse rates and duration of illness (43) (44) (45) (46) (8) (47), meaning the algorithms should apply generally to people with MS, rather than just to specific sub-groups. In addition, the work undertaken to explore the content overlap between the measures provided a form of 'triangulation' in assessing the appropriateness of the mapping algorithms. Drawing on good quality qualitative research findings regarding the impacts of fatigue on HRQoL and developing a conceptual framework, provided unique insights into why the measures did and did not map well.

It is acknowledged that mapping methods are a second-best option to directly collected HSUVs for estimating QALYs (29) (48) (41). Use of mapping increases the uncertainty and error around estimates of HSUVs (29), and is particularly problematic when there is little content overlap or relationship between the measures being mapped to and from (41). However, when PBM data are not collected directly in a trial, empirically-evidenced mapping algorithms may be used. With the exception of the EQ-5D, the algorithms reported here can be used to support improvements in decision-making where primary PBM data are unavailable.

Conclusions

We present statistical algorithms that allow data from the FSS, a fatigue-specific patient-reported outcome measure, to be used in the estimation of QALYs, which are a suitable and policy-relevant measure for use in cost-effectiveness analyses. This will enable the results of studies using the FSS to inform decision-making in a health technology assessment context.

Declarations

Ethics approval and consent to participate

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

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from SWIMS Data-Sharing Committee.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Multiple Sclerosis Society of Great Britain and Northern Ireland and the UK NIHR Collaboration for Leadership in Applied Health Research and Care of the South West Peninsula (PenCLAHRC). The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

The views expressed in this publication are those of the authors and not necessarily those of the Multiple Sclerosis Society, the UK NIHR or the Department of Health.

The Multiple Sclerosis Society of Great Britain and Northern Ireland and the Peninsula Medical School Foundation provided support for the SWIMS Project.

Authors' contributions

All authors conceived the idea for the research, EG conducted the data analysis with support and supervision from AH, EG drafted the article, and CG and AH provided suggestions/edits etc, all authors approved the final version of the paper.

Acknowledgements

The authors are grateful to the SWIMS Project participants for allowing access to data they provided for the SWIMS Project. The authors acknowledge the SWIMS Project Team for delivering these data. This publication is the work of the authors, who will serve as guarantors for the contents of this publication. This publication does not necessarily reflect the views of the SWIMS Project Team nor the SWIMS Data-Sharing Committee.

List of abbreviations:

CLAD	Censored least absolute deviation
EQ-5D	EuroQoL EQ-5D-3L
FSS	Fatigue Severity Scale
HSUV	health state utility value
MS	multiple Sclerosis
MSIS-8D	Multiple Sclerosis Impact Scale-8D
NICE	National Institute for Health and Care Excellence
OLS	Ordinary least squares
PBM	preference-based measure

QALY	quality-adjusted life-year
SF-6D	Short-Form 6D
SWIMS	South West Impact of MS study
TTO	Time trade-off

References

1. Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2007.
2. Zajicek J, Freeman J, Porter B. Multiple Sclerosis Care: A Practical Manual. Oxford: Oxford University Press; 2007.
3. Flachenecker P, Kümpfel T, Kallmann B, Gottschalk M, Grauer O, Rieckmann P, et al. Fatigue in multiple sclerosis: a comparison of different rating scales and correlation to clinical parameters. Multiple Sclerosis. 2002;8:523-6.
4. Tomassini V, Pozzilli C, Onesti E, Pasqualetti P, Marinelli F, Pisani A, et al. Comparison of the effects of acetyl l-carnitine and amantadine for the treatment of fatigue in multiple sclerosis: results of a pilot, randomised, double-blind, crossover trial. Journal of the Neurological Sciences. 2004;218:103-8.
5. Shaygannejad V, Janghorbani M, Ashtari F, Zakeri H. Comparison of the effect of aspirin and amantadine for the treatment of fatigue in multiple sclerosis: a randomized, blinded, crossover study. Neurological Research. 2012;34:854-8.
6. Rammohan K, Rosenberg J, Lynn D, Blumenfeld A, Pollak C, Nagaraja H. Efficacy and safety of modafinil (Provigil®) for the treatment of fatigue in multiple sclerosis: a two centre phase 2 study. Journal of Neurology, Neurosurgery and Psychiatry. 2002;72:179–83.
7. van Kessel K, Moss-Morris R, Willoughby E, Chalder T, Johnson M, Robinson E. A Randomized Controlled Trial of Cognitive Behavior Therapy for Multiple Sclerosis Fatigue. Psychosomatic Medicine. 2008;70:205-13.
8. Jones K, Ford D, Jones P, John A, Middleton R, Lockhart-Jones H, et al. How people with multiple sclerosis rate their quality of life: an EQ-5D survey via the UK MS Register. PLoS ONE. 2013;8(6):e65640.
9. Guidelines for the economic evaluation of health technologies 3rd edition. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH); 2006.
10. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3). Barton, Australia: Pharmaceutical Benefits Advisory Committee, Australian Government, Department of Health and Ageing; 2008.
11. Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, et al. The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. Pharmacoeconomics. 2015;33:993–1011.
12. Hawton A, Green C, Telford C, Zajicek J, Wright D. Using the Multiple Sclerosis Impact Scale to Estimate Health State Utility Values: Mapping from the MSIS-29, Version 2, to the EQ-5D and the SF-6D. Value in Health. 2012;15:1084-91.
13. Versteegh M, Rowen D, Luime J, Boggild M, Groot CU-d, Stolk E. Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. Medical Decision Making. 2012;32:554–68.
14. Hawton A, Green C, Telford C, Wright D, Zajicek J. The use of multiple sclerosis condition-specific measures to inform health policy decision-making: mapping from the MSWS-12 to the EQ-5D. Multiple Sclerosis. 2012;18:853–61.
15. Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence; 2013.
16. Wailoo A, Hernandez-Alava M, al AMe. Mapping to Estimate Health-State Utility from Non–Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. Value in Health. 2017;20(1):18-27.
17. Krupp L, LaRocca N, Muir-Nash J, Steinberg A. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Archives of Neurology. 1989;46:1121-23.
18. Learmonth Y, Dlugonski D, Pilutti L, Sandroff B, Klaren R, Motl R. Psychometric properties of the Fatigue Severity Scale and the Modified Fatigue Impact Scale. Journal of the Neurological Sciences. 2013;331:102–7.

19. Valko P, Bassetti C, Bloch K, Held U, Baumann C. Validation of the Fatigue Severity Scale in a Swiss Cohort. Sleep. 2008;31(11):1601-7.
20. Armutlu K, Korkmaz N, Keser I, Sumbuloglu VA, DI, Guney Z, Karabudak R. The validity and reliability of the Fatigue Severity Scale in Turkish multiple sclerosis patients. Int J Rehabil Res. 2007;30:81-5.
21. Hjollund N, Andersen J, Bech P. Assessment of fatigue in chronic disease: a bibliographic study of fatigue measurement scales. Health Qual Life Outcomes. 2007;5(12).
22. Dolan P. Modeling Valuations for EuroQol Health States. Medical Care. 1997;35:1095-108.
23. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics. 2002;21:271-92.
24. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. J Epidemiol Community Health. 1999;53:46–50.
25. 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 in Health. 2015;18:1016-24.
26. Goodwin E, Green C, Spencer A. Estimating a preference-based index for an eight dimensional health state classification system derived from the Multiple Sclerosis Impact Scale (MSIS-29). Value in Health. 2015;18:1025-36.
27. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technology Assessment. 2009;13(12).
28. 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.
29. Longworth L, Rowen D. Technical Support Document 10: The use of mapping methods to estimate health state utility values. National Institute for Health and Care Excellence Decision Support Unit; 2011.
30. Riazi A. Patient-reported Outcome Measures in Multiple Sclerosis. The International MS Journal. 2006;13:92–9.
31. Ware J. Conceptualization and measurement of health-related quality of life: comments on an evolving field. Archives of Physical Medicine and Rehabilitation. 2003;84(Suppl 2):S43-S51.
32. European Medicines Agency. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. EMEA/CHMP/EWP/139391/2004. London: European Medicines Agency; 2005.
33. Nunnally J, Bernstein I. Psychometric Theory. New York: McGraw-Hill; 1994.
34. Cohen J. Statistical Power Analysis for the Behavioural Sciences. Hillsdale NJ: Lawrence Erlbaum Associates; 1988.
35. Brazier J, Yang Y, suchiya T, 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.
36. Powell J. Least absolute deviations estimation for the censored regression model. Journal of Econometrics. 1984;25:303–25.
37. HERC database of mapping studies Version 7.0. http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies 24th April 2019 [
38. Tremmas I, Petsatodis G, Potoupnis M, Laskou S, Giannakidis D, Mantalovas S, et al. Monitoring changes in quality of life in patients with lung cancer under treatment with chemotherapy and co administration of zoledronic acid by using specialized questionnaires. Journal of Cancer. 2018;9(10):1731-6.
39. Rosa K, Fu M, Gilles L, Cerri K, Peeters M, Bubb J, et al. Validation of the Fatigue Severity Scale in chronic hepatitis C. Health & Quality of Life Outcomes. 2014;12(90).

40. Ernstsson O, Tingho P, Alexanderson K, Hillert J, Burstro¨m K. The External Validity of Mapping MSIS-29 on EQ-5D Among Individuals With Multiple Sclerosis in Sweden. MDM Policy & Practice. 2017;2:1-9.
41. Round J, Hawton A. Statistical Alchemy: Conceptual Validity and Mapping to Generate Health State Utility Values. PharmacoEconomics Open. 2017;1(4):233-9.
42. Sidovar M, Limone B, Lee S, Coleman C. Mapping the 12-item multiple sclerosis walking scale to the EuroQol 5-dimension index measure in North American multiple sclerosis patients. BMJ Open. 2013;3.
43. Confavreaux C, Compston A. The natural history of multiple sclerosis. In: Compston A, editor. McAlpine's Multiple Sclerosis. Philadelphia: Churchill Livingstone Elsevier; 2006.
44. Ford H, Gerry E, Airey C, Al E. The prevalence of multiple sclerosis in the Leeds Health Authority. Journal of Neurology, Neurosurgery and Psychiatry. 1998;64:605-10.
45. Forbes R, Wilson S, Swingler R. The prevalence of multiple sclerosis in Tayside, Scotland: do latitudinal gradients really exist? Journal of Neurology, Neurosurgery and Psychiatry. 1999;246:1033-40.
46. Fox C, Bensa S, Bray I, Zajicek J. The epidemiology of multiple sclerosis in Devon: a comparison of new and old classification criteria. Journal of Neurology, Neurosurgery and Psychiatry. 2004;75:56-60.
47. Robertson N, Deans J, Fraser M, Al E. Multiple sclerosis in the North Cambridgeshire districts of East Anglia. Journal of Neurology, Neurosurgery and Psychiatry. 1995;59:71-6.
48. McCabe C, Edlin R, Meads D, Brown C, Kharroubi S. Constructing indirect utility models: some observations on the principles and practice of mapping to obtain health state utilities. Pharmacoeconomics. 2013;31(8):635-41.

Table 1: Summary of respondent characteristics, comparison of estimation and validation datasets

	All baseline data			Estimation dataset			Validation dataset			Difference[1]	
	Mean	**SD**	**Observations**	**Mean**	**SD**	**Observations**	**Mean**	**SD**	**Observations**	**t statistic**	**p value**
Measure											
FSS	43.73	15.10	1054	43.44	15.16	787	44.60	14.91	267	-1.085	0.278
EQ-5D	0.596	0.295	1346	0.600	0.291	1005	0.584	0.309	341	0.831	0.406
SF-6D	0.646	0.130	632	0.650	0.135	473	0.636	0.113	159	1.141	0.254
MSIS-8D	0.646	0.185	690	0.647	0.190	523	0.641	0.172	167	0.412	0.681
Characteristic											
Age	50.67	11.68	1400	50.74	11.73	1048	50.45	11.54	352	0.402	0.688
Duration (years)	9.62	10.01	1347	9.61	10.00	1009	9.68	10.09	338	-0.113	0.910
EDSS score	4.30	2.31	289	4.32	2.34	218	4.22	2.24	71	0.324	0.746
	Percentage		**Observations**	**Percentage**		**Observations**	**Percentage**		**Observations**	**chi^2 statistic**	**p value**
Gender											
Female	73.86%		1040	74.62%		788	71.59%		252	1.256	0.262
Male	26.14%		368	25.38%		268	28.41%		100		
MS type											
Relapsing remitting	41.97%		572	42.66%		439	39.82%		133	7.572	0.109
Primary progressive	19.37%		264	18.56%		191	21.86%		73		
Secondary progressive	16.95%		231	17.69%		182	14.67%		49		
Benign	3.3%		45	3.69%		38	2.10%		7		
DK or combination	18.42%		251	17.40%		179	21.56%		72		
Missing			45			27			18		
Recent relapse[2]											
Yes	53.55%		732	53.42%		546	53.91%		186	0.025	0.988
No	33.28%		455	33.37%		341	33.04%		114		
Don't know	13.17%		180	13.21%		135	13.04%		45		
Missing			41			34			7		

SD = standard deviation; FSS = Fatigue Severity Scale; EQ-5D = EuroQoL EQ-5D-3L; SF-6D = Short-Form 6D; MSIS-8D = Multiple Sclerosis Impact Scale – Eight Dimensions; EDSS = Expanded Disability Status Scale; DK = don't know.

[1]Difference between estimation and validation datasets

[2]relapse in the 12 months prior to completing the baseline questionnaire

nb response numbers are lower for the SF-6D and the MSIS-8D as only version 2 of these questionnaires were included

Table 2: Comparison of measures against conceptual framework

Conceptual framework	Fatigue severity scale	EQ-5D	SF-6D	MSIS-8D
Descriptions of fatigue				
General fatigue or vitality	3. Easily fatigued 5. Causes frequent problems 8. Among most disabling symptoms	-	6. Vitality	-
Physical effects				
General	4. Interferes with physical functioning 6. Prevents sustained physical functioning	-	1. Physical functioning	1. Physically demanding tasks
Triggers		-	-	-
Specific physical effects	2. Exercise brings on fatigue	4. Pain/Discomfort 1. Mobility	4. Pain	3. Being stuck at home
Mental effects				
General	-	-	-	5. Feeling mentally fatigued
Specific psychological effects	1. Motivation is lower	5. Anxiety/Depression	5. Mental health	6. Irritable, impatient, short-tempered 8. Feeling depressed
Specific cognitive effects	-	-	-	7. Problems concentrating
Indirect effects	-	-	-	-
Participation effects				
General	7. Interferes with duties & responsibilities 9. Interferes with work, family, social life		-	3. Being stuck at home
Effects on specific activities		2. Self-Care 3. Usual Activities	1. Physical functioning 2. Role limitations 3. Social functioning 4. Pain	2. Social and leisure activities 4. Work or other daily activities

EQ-5D = EuroQoL EQ-5D-3L; SF-6D = Short-Form 6D; MSIS-8D = Multiple Sclerosis Impact Scale – Eight Dimensions

Explanation for allocation of particular items:

- SF-6D Physical functioning: included under both “Physical effects” and “Functioning/ participation” because level descriptions include “moderate/ vigorous activities” and “bathing and dressing”
- SF-6D Pain: included under both “Physical effects” and “Functioning/ participation” because level descriptions include “pain that interferes with your normal work”
- SF-6D Mental health: included under “Specific psychological effects” because level descriptions refer to feeling “tense or downhearted and low”
- SF-6D Role limitations: included under “Functioning/ participation – activities” because level descriptions refer to “work or other regular daily activities”.
- MSIS-8D Being stuck at home: included under “Specific physical effects” because the MSIS-8D uses this question as a proxy for mobility, however we have also included it here under “Functioning/ participation”

Table 3: Correlations between Fatigue Severity Scale and preference-based measures

FSS total score and PBM dimensions	rho	Observations
EQ-5D versus FSS total score		
Mobility	0.423	1035
Self-care	0.385	1048
Usual activities	0.524	1051
Pain/Discomfort	0.361	1047
Anxiety/Depression	0.292	1049
SF-6D versus FSS total score		
Physical functioning	0.547	649
Role limitations	0.424	645
Social functioning	0.530	644
Pain	0.429	642
Mental health	0.324	648
Vitality	0.615	654
MSIS-8D versus FSS total score		
Physically demanding tasks	0.585	656
Social and leisure activities	0.560	652
Mobility (being stuck at home)	0.489	656
Work or other daily activities	0.558	655
Feeling mentally fatigued	0.582	656
Feeling irritable, impatient or short-tempered	0.377	654
Problems concentrating	0.450	654
Feeling depressed	0.320	653
EQ-5D versus FSS item		
1 My motivation is lower	-0.285	1040
2 Exercise brings on my fatigue	-0.382	1038
3 I am easily fatigued	-0.464	1040
4 Interferes with physical functioning	-0.471	1033
5 Causes frequent problems for me	-0.498	1039
6 Prevents sustained physical functioning	-0.527	1040
7 interferes with duties and responsibilities	-0.536	1038
8 Among my most disabling symptoms	-0.336	1035
9 Interferes with work, family or social life	-0.482	1039
SF-6D versus FSS item		
1 My motivation is lower	-0.400	614
2 Exercise brings on my fatigue	-0.409	614
3 I am easily fatigued	-0.545	614
4 Interferes with physical functioning	-0.541	612
5 Causes frequent problems for me	-0.585	614
6 Prevents sustained physical functioning	-0.575	614
7 interferes with duties and responsibilities	-0.623	613
8 Among my most disabling symptoms	-0.455	610
9 Interferes with work, family or social life	-0.603	614
MSIS-8D versus FSS item		
1 My motivation is lower	-0.387	659
2 Exercise brings on my fatigue	-0.423	659
3 I am easily fatigued	-0.560	659

4 Interferes with physical functioning	-0.554	656
5 Causes frequent problems for me	-0.615	659
6 Prevents sustained physical functioning	-0.606	660
7 interferes with duties and responsibilities	-0.637	660
8 Among my most disabling symptoms	-0.428	656
9 Interferes with work, family or social life	-0.596	659

All coefficients significant at p<0.0001
PBM = preference-based measure; FSS = Fatigue Severity Scale; EQ-5D = EuroQoL EQ-5D-3L; SF-6D = Short-Form 6D; MSIS-8D = Multiple Sclerosis Impact Scale – Eight Dimensions

Table 4: Discriminative validity

EQ-5D vs FSS groups	Mean	SD	Obs	SES
Mild/no fatigue	0.775	0.218	297	0.615
Moderate fatigue	0.641	0.233	369	0.803
Severe fatigue	0.454	0.3	357	
FFS total	0.614	0.285	1,023	
F-statistic	131.84	Prob <	0.0001	
Bartlett's chi2	40.065	Prob <	0.0001	
SF-6D vs FSS groups				
Mild/no fatigue	0.747	0.124	189	0.871
Moderate fatigue	0.639	0.099	225	0.879
Severe fatigue	0.552	0.083	193	
FFS total	0.645	0.129	607	
F-statistic	172.46	Prob <	0.0001	
Bartlett's chi2	30.047	Prob <	0.0001	
MSIS-8D vs FSS groups				
Mild/no fatigue	0.764	0.115	202	0.739
Moderate fatigue	0.679	0.134	240	1.381
Severe fatigue	0.494	0.186	208	
FFS total	0.646	0.184	650	
F-statistic	180.71	Prob <	0.0001	
Bartlett chi2	51.434	Prob <	0.0001	

SD = standard deviation; obs = observations; SES = standardised effect size; FSS = Fatigue Severity Scale; EQ-5D = EuroQoL EQ-5D-3L; SF-6D = Short-Form 6D; MSIS-8D = Multiple Sclerosis Impact Scale – Eight Dimensions

Table 5: Models mapping from FSS total to PBMs using estimation dataset

	EQ-5D				SF-6D						MSIS-8D			
	CLAD A		CLAD C		OLS A		CLAD A		CLAD C		CLAD A		CLAD C	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
FSS total	-0.006*	0.0006	-0.006*	0.0006	-0.006*	0.0003	-0.006*	0.0004	-0.006*	0.0004	-0.007*	0.0007	-0.008*	0.0008
Age			-0.003*	0.0007					-0.0005	0.0005			-0.001	0.0008
Female			0.012	0.0133					-0.012	0.0107			-0.024	0.0233
Constant	0.921	0.0256	1.058	0.0610	0.897	0.0151	0.913	0.0195	0.966	0.0374	0.985	0.0228	1.084	0.0719
Observations	763		755		455		455		452		474		464	
F statistic					357.45									
Prob>F					<0.0001									
R-squared					0.451									
Pseudo R2	0.107		0.126				0.267		0.274		0.196		0.194	
Coefficients	1		3		1		1		3		1		3	
Significant coefficients	1		2		1		1		1		1		1	
Mean absolute error (MAE)	0.175		0.173		0.078		0.078		0.077		0.117		0.116	
Mean squared error (MSE)	0.066		0.067		0.01		0.01		0.01		0.024		0.023	
Root MSE	0.257		0.258		0.1		0.1		0.1		0.154		0.152	
Normalised root MSE	16.12%		16.19%		14.31%		14.31%		14.31%		19.18%		18.93%	
Individuals with MAE < 0.25	78.37%		79.34%		98.68%		98.68%		98.45%		89.05%		90.41%	
Individuals with MAE < 0.1	47.05%		49.14%		68.13%		69.01%		70.13%		51.93%		51.84%	
Individuals with MAE < 0.05	26.47%		29.14%		41.32%		41.98%		42.48%		28.40%		29.39%	

*p<0.001

Coeff = model coefficient; SE = standard error; CLAD = Censored Least Adjusted Deviation model; EQ-5D = EuroQoL EQ-5D-3L; SF-6D = Short-Form 6D; MSIS-8D = Multiple Sclerosis Impact Scale – Eight Dimensions

Table 6: Models mapping from FSS total to PBMs using validation dataset

	EQ-5D				SF-6D						MSIS-8D			
	CLAD A		CLAD C		OLS A		CLAD A		CLAD C		CLAD A		CLAD C	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
FSS total	-0.007*	0.0012	-0.008*	0.0011	-0.004*	0.0005	-0.004*	0.0007	-0.004*	0.0008	-0.006*	0.0010	-0.006*	0.0011
Age			-0.004*	0.0011					0.0004	0.0009			-0.001	0.0020
Female			-0.009	0.0260					0.002	0.0187			0.012	0.0395
Constant	1.001	0.0549	1.233	0.0979	0.81	0.0261	0.793	0.0394	0.827	0.0781	0.939	0.0432	0.974	0.1252
Observations	260		258		152		152		152		157		157	
F statistic					54.71								0.185	
Prob>F					<0.0001									
R-squared					0.316									
Pseudo R2	0.119		0.141				0.169		0.169					
Coefficients	1		3		1		1		3		1		3	
Significant coefficients	1		2		1		1		1		1		1	
Mean absolute error (MAE)	0.183		0.179		0.068		0.068		0.071		0.118		0.114	
Mean squared error (MSE)	0.076		0.071		0.008		0.008		0.009		0.023		0.022	
Root MSE	0.276		0.267		0.09		0.09		0.095		0.151		0.149	
Normalised root MSE	17.31%		16.75%		12.88%		12.88%		13.59%		18.80%		18.56%	
Individuals with MAE < 0.25	78.85%		76.92%		98.68%		98.03%		98.03%		92.36%		91.08%	
Individuals with MAE < 0.1	49.62%		47.31%		76.32%		75.00%		75.66%		50.32%		52.23%	
Individuals with MAE < 0.05	24.62%		25.77%		48.68%		46.05%		46.05%		22.93%		31.21%	

*p<0.001

Coeff = model coefficient; SE = standard error; CLAD = Censored Least Adjusted Deviation model; EQ-5D = EuroQoL EQ-5D-3L; SF-6D = Short-Form 6D; MSIS-8D = Multiple Sclerosis Impact Scale – Eight Dimensions

Table 7: Mean absolute errors by severity group

FSS to EQ-5D	**CLAD Model A**	**CLAD Model C**	**OLS Model A**
EQ_5D<=0.65	0.234	0.238	
EQ_5D>0.65	0.123	0.115	
FSS to SF-6D			
SF_6D<=0.65	0.070	0.070	0.070
SF_6D>0.65	0.088	0.088	0.088
FSS to MSIS-8D			
MSIS_8D<=0.7	0.154	0.154	
MSIS_8D>0.7	0.082	0.082	
Cut-off points for EQ-5D, SF-6D and MSIS-8D were chosen to give roughly equally-sized groups.			

Additional file 1

Development of a conceptual framework describing the impact of fatigue on people with MS: a systematic review of the literature

Aim: to identify the main impacts of fatigue on the quality of life of people with MS, from the perspective of people with MS.

Objective: to review the qualitative literature on the impact of fatigue on the lived experiences of people with MS

Methods

Literature search methods

A search design was developed based on three key components of the objective of the literature review: multiple sclerosis, fatigue and qualitative methods.

MS search terms were based on those used for routine searches undertaken by the Cochrane Collaboration's "Multiple sclerosis and rare diseases of the central nervous system" group [Cochrane 2017].

Fatigue search terms were based on those used in a review of interventions for fatigue in Parkinson's disease undertaken by the Cochrane Movement Disorders Group [Elbers et al, 2014].

Qualitative search terms were based on those developed for the purposes of a study that investigated how to find qualitative research in the context of the medical literature [Shaw et al, 2004].

Search terms within each component were combined using the Bayesian operator "or". The components were combined using the "and" operator.

Inclusion criteria

- Original research using a qualitative methodology
- Participants are people with MS, or include people with MS alongside people with other conditions, where the results for people with MS are separately identifiable
- Papers with a stated aim of investigating the impact of fatigue on one or more aspects of (health-related) quality of life, well-being, functioning or participation
- English language

Exclusion criteria

- Review papers
- Papers that explore one or more aspects of (health-related) quality of life, well-being, functioning or participation in MS, without an *a priori* focus on fatigue.
- Papers that focused on fatigue, but did not report on the impact of fatigue on (health-related) quality of life, well-being, functioning or participation.

- Papers that did not present separately identifiable results for people with MS.

Analysis methods

The purpose of this review is to provide the background for mapping between the FSS and the EQ-5D, SF-6D and MSIS-8D. The measurement concept underpinning the three selected PBMs is health-related quality of life. Therefore, the results of the literature review were analysed using a conceptual framework based on the key dimensions of health-related quality of life.

There is no firm agreement on which dimensions comprise HRQoL, however there is a consensus that, at a minimum, physical and psychological domains should be included [Riazi, 2006]. More recently, a third dimension, relating to social and role function and participation (ie one's ability to perform 'normal' or expected activities and roles), has been added [Ware, 2003]. Therefore, the conceptual framework consisted of three broad domains: physical functioning, psychological and cognitive functioning, and social functioning and participation.

The themes that were identified by the original analysis were extracted from the results sections of the papers and pasted into tables, including the name of the theme and a description of its contents. All identified themes were then allocated to the three domains of HRQoL. Any themes that were repeated across more than one study were combined and any links between themes were noted. Themes that did not fit into the HRQoL domains were collated separately.

In the next stage, sub- domains were developed by grouping together themes that described similar concepts within each domains of HRQoL. The themes that did not fit into the HRQoL domains and the links between themes were explored to determine whether amendments needed to be made to the three-dimensional structure of the conceptual framework. This was then used to produce a conceptual framework of how fatigue affects the HRQoL of people with MS, for use in assessing the content validity of the source and target measures.

Results

Literature search results

The literature search returned 1124 results. Based on the titles and abstracts of these, 1062 were excluded from further consideration. Of the remaining 62 studies, 11 were conference abstracts for which the full text was not available. The full text of the remaining 52 papers was obtained, and these were assessed against the inclusion and exclusion criteria. Seventeen papers were excluded because they focussed on aspects of fatigue other than its impact (n=7), they did not focus on fatigue (6), they did not use qualitative methods (3) or they were not primarily concerned with MS. Therefore, twelve papers remained for inclusion in the review. This is summarised in Figure A1.

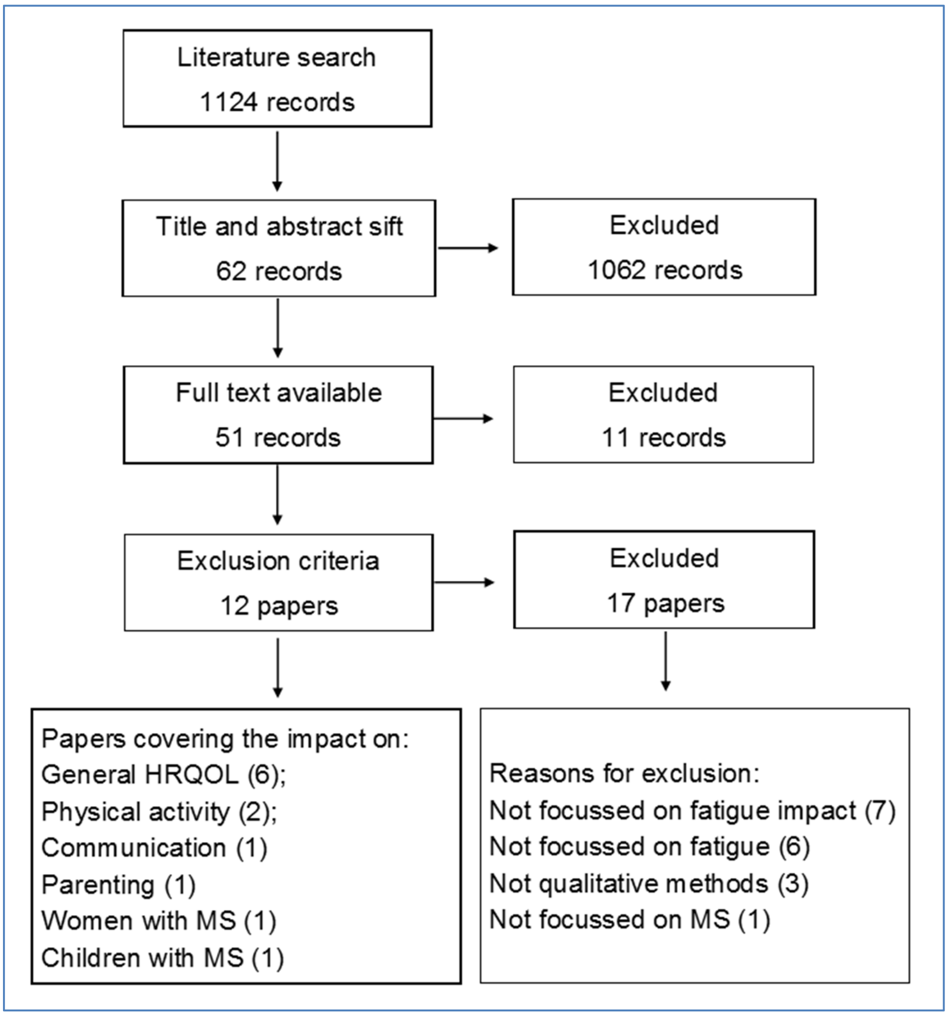


Figure A1: Literature search results

Development of the conceptual framework

Most of the themes that had been identified in the original qualitative research studies fitted into the three domains of HRQoL that were defined *a priori*. There were two notable exceptions. Several of the themes described the experience of fatigue itself, rather than its effect on HRQoL. This experience was clearly of great important to the people with MS who contributed to the original research, and underpinned the ways in which fatigue impacts upon HRQoL. Therefore, an additional domain was added: "Descriptions of fatigue". In terms of the links between themes, a clear relationship emerged between "functioning and participation" and "psychological well-being". People with MS specifically identified negative effects on their psychological well-being that were caused by the impact of their fatigue on their functioning and participation. These stood alongside, but distinct from, the direct impact of fatigue on psychological well-being. Therefore, this became a domain in its own right.

Tables A1 – A4 outline how the themes identified from the literature were mapped to the conceptual framework. The conceptual framework is illustrated in Figure A2.

Table A1. Descriptions of fatigue

1.1 Fatigue as a whole body experience	Fatigue was experienced in the muscles, head, and entire body. It affects sensation in the whole body, from the hair to the toes.
	Fatigue was perceived in the body – whole of parts of the body were perceived differently than they were before. The body did not feel natural and could not be taken for granted; increased awareness of the body all the time. Strained body with diminished power.
	Two opposite perceptions of the body: (1) heavy & painful (2) numbed, dead, not quite awake, as if parts were missing.
	A feeling of having a heavy body; wanting to let their arms and hands hang down; impossible to raise the arms or hold the body up straight. Muscles feel too weak to support the body
1.2 Betrayed by your own body	Some felt betrayed when fatigue invaded the body. The body was hard to control and couldn't be trusted.
	Body will not obey, eg try to lift leg and nothing happens.
	Their own bodies ruled them and they had to adjust themselves. Feeling feeble and unable to manage.
1.3 All-consuming fatigue	Participants experienced fatigue much of the time, and when they did not, they were thinking about it – always taking it into account.
	An ever-present, ongoing experience, unrelenting and virtually ever present, even after rest or sleep.
	A paralyzing force; although some small reserve of energy is still available, they feel virtually powerless to perform desired activities.
	Undertow Effect: suffocating fatigue characterized by energy impoverishment and absolute powerlessness, relieved only by sleep.
	Energy loss was very unpleasant, perceived as a form of paralyses and as an unstoppable destructive force invading the body, leaving participants unable to manage anything further.
1.4 An unusual and invisible feeling	A unique and novel sensation, an experience that is different from experiences of being tired when healthy
	Fatigue is invisible and difficult to describe
1.5 Characteristics of the experience of fatigue	Weariness, sleepiness, tired most days, weak at rest, exhausted after minimal activity
	Sudden, can happen very rapidly, unpredictable, uncontrollable
	Sensation of one's batteries running out
	Need day rest or sleep
	Unrefreshing or broken nocturnal sleep
1.6 Interactions with other symptoms	Interaction of fatigue with other symptoms leads to difficulties
	Fatigue that worsens along with other symptoms
	Fatigue can exacerbate other MS symptoms and vice versa
Individuals were affected differently by fatigue, and could experience fatigue in one or more different ways during the course of their MS.	
It was often hard to differentiate whether a participant was discussing fatigue or MS, as these terms seemed to be used interchangeably	
Many participants had experienced more than one state of fatigue during the course of their MS and, on occasion, one state of fatigue could trigger another.	
Rather than isolating fatigue, it's more about the complex and unpredictable relationship between fatigue and other symptoms.	

Table A2. Physical effects of fatigue

2.1 Specific physical effects	Difficulty walking
	Falling over
	Weakness/ muscle weakness
	Participants described different states of fatigue.
	Limbs heavy
	Speech problems
	Coordination
	Sensory disturbances/ visual disturbances
	Flickering and swimming or pounding sensation in their eyes, causing dizziness and nausea.
	Pain
	Balance.
	A temporary increase in physical symptoms was associated with increased fatigue.
2.2 Physical triggers	Physical exertion induces weakness/ worsens fatigue
	Fatigue due to unexpected actions
	The feeling of being fatigued increased because of the extra effort of arranging footsteps when walking.

Table A3. Mental effects of fatigue

3.1 Psychological effects of fatigue	Emotional impact of fatigue	
	A temporary increase in emotional symptoms was associated with increased fatigue.	
	Participants described different states of fatigue including feelings of depression.	
	Feelings of defeat	
3.2 Cognitive effects of fatigue	General	Cognitive impact of fatigue
		Participants described different states of fatigue including mental fogginess.
		Felt that their brain was not totally clear; felt like being struck in the head by a sledgehammer.
		Head experiences: “Brain-cheese,” a “hazy, out-of-body fatigue feeling,” and a “hangover.”
		Links to other symptoms, including cognition.
		A temporary increase in cognitive symptoms was associated with increased fatigue.
		Rather than isolating fatigue, it’s about the complex and unpredictable relationship between fatigue and cognition.
	Specific	Difficulty concentrating or thinking clearly
		Perception of lower cognitive ability and energy
		Impact on daily life: difficulties in making decisions and plans
		Difficulty solving complex problems

		Difficulty withstanding disturbing sounds
		Difficulties in remembering
		Making mistakes
		Only for brief moments could they feel totally focused.
		Not being able to look forward in time, thinking in the present moment.
3.3 Psychological triggers of fatigue	Cognitive	Mental effort worsens fatigue
	Emotional	Fatigue due to a change in mood
		Stress/anxiety worsens fatigue
		Vicious circle: thinking/ worrying about fatigue could cause fatigue, leaving them unable to complete the task
3.4 Indirect psychological effects of fatigue **(links to participation and functioning effects)**	Emotional impacts	A feeling of having the will but not the ability: wanted to live life as before and be an active person.
		Anxiety
		Helpless and exposed.
		Insecurity
		Frustration, stress, sadness.
		Dissatisfaction
		Lower self-worth, despair, sorrow
		Shame; being misunderstood (eg being mistaken as being drunk).
		Anger
	Enjoyment of life	Involuntary isolation
		Inability to enjoy social activities or hobbies
		No fun in life, feeling bored.
		Feeling trapped by having to live a very structured life; loss of spontaneity
		Prevention of a “normal” life due to fatigue
		Forced interruption of activities due to fatigue
		Inadequate satisfaction of one’s basic needs
	Identity	Loss of sense of self due to fatigue
		Disappointment in a fatigued self
		Inability to tend to appearances due to fatigue
		Non-achievement of goals due to a gap in the expected and actual behavioral potential
		Loss of control, which appeared to threaten the self-integrity of the individual.
		Progressive losses including work, youth, driving, strength and energy, relationship roles; feeling “old before my time.” Losses of driver’s license and employment had emotional effects and challenged men’s self-identities within the family. Some felt they had progressively lost strength and energy, attributes they linked to “being a man.” Some described attributes associated with self-identity that either contributed to fatigue or helped them continue their exercise despite fatigue, ie stubbornness, determination.

	Those who were able to engage in valued activities—even if the intensity was less, or if they achieved them through a different route—experienced positive feelings and a sense of control. Those who were unable to make goal adjustments disengaged from valued activities, resulting in negative feelings.
	Ongoing frustration can lead to depression, particularly "when the frustrated goal is deeply connected to the core of the self".
	'The importance of having goals that were highly valued and related to activities and work prior to diagnosis allowed the men to feel a sense of achievement and optimism despite their losses.'

Table 4. Effects of fatigue on participation and functioning

4.1 Pervasive impact	Influences all activities and responsibilities at work, home, and play.	
	Restrictions or interruptions to life, including changes in roles within the family, social life and working situation.	
	Barriers to participation were not perceived to directly result from any single MS impairment eg fatigue or communication, but from a complex interplay between the impairments experienced by an individual, the coping strategies employed and people's attitudes.	
4.2 Activities	Put things off, force self to do things	
	Unable to carry out daily tasks as could before	
	Activities of daily living	
	Housework	
	Giving up work, working fewer hours	
	Decreased opportunities for social interaction.	
	Social activities/ hobbies	
4.3 Effects of strategies to manage fatigue	Implications of having to plan ahead/ lead structured daily life/ build rest periods into daily routine = less opportunity for spontaneity	
	Implications of having to reduce overall activity or prioritise certain activities over others = dilemmas over which things don't get done	
	Implications of having to take a planned or necessary cessation of physical activity	
	Difficulty of employment due to the measures for treating fatigue, related to an interruption of activities	
	May take up formal exercise, or other physical activities, in attempt to enhance resistance to fatigue	
	Time-consuming: can't hurry, need to take time and avoid stressful situations; doing things in advance, calmly and methodically.	
4.4 Roles and relationships	Communication and fatigue	Difficult for others to understand the person's experiences and needs because fatigue is "invisible" and difficult to describe
		Fatigue increases the frequency and severity of communication symptoms, language-processing deficits, motor speech symptoms**
		Some communication symptoms occur only when experiencing fatigue - language processing difficulties and dysarthria**
		The interplay between fatigue and communication led to communication symptoms becoming more apparent to listeners. The resulting communication did not reflect how they would like to represent themselves (eg drunk or lazy rather than able and competent).**

		Common to all participants was the enormous effort and pre-planning that remains hidden from communication partners, eg dealing with word finding and memory difficulties, keeping interactions operating as normally as possible on the surface.**
	Handling fatigue in relation to others	Concealment, eg measures to limit activity without letting others know
		Measures to arrange an environment by gaining the support of others
	A feeling of being absent	They felt as if they had been split in two parts: one part was participating while the other was just watching. Feeling both present and absent: seeing everything but feeling as if they weren't there. Feeling anaesthetized; things just passing by.
		Unable to understand things happening around them or to participate in conversations due to lack of concentration.
	Letting people down/ causing problems	Feeling unreliable and could not always keep promises. Leaving everything half-done due to unpredictable fatigue.
		Unable to participate in family activities; felt this was difficult for the rest of the family - the whole family was suffering.
		Problems in one's life and friendships due to unpredictable fatigue
		Concern of causing friends trouble due to fatigue
	Dependency	Perceptions of dependency - trapped in the sense of needing help from other people – involves feelings of being a burden
4.5 Participation triggers	Trying to accomplish too much	
	Family, work or socioeconomic stress	
	Continuous nature of burdens and actions	
	Work	

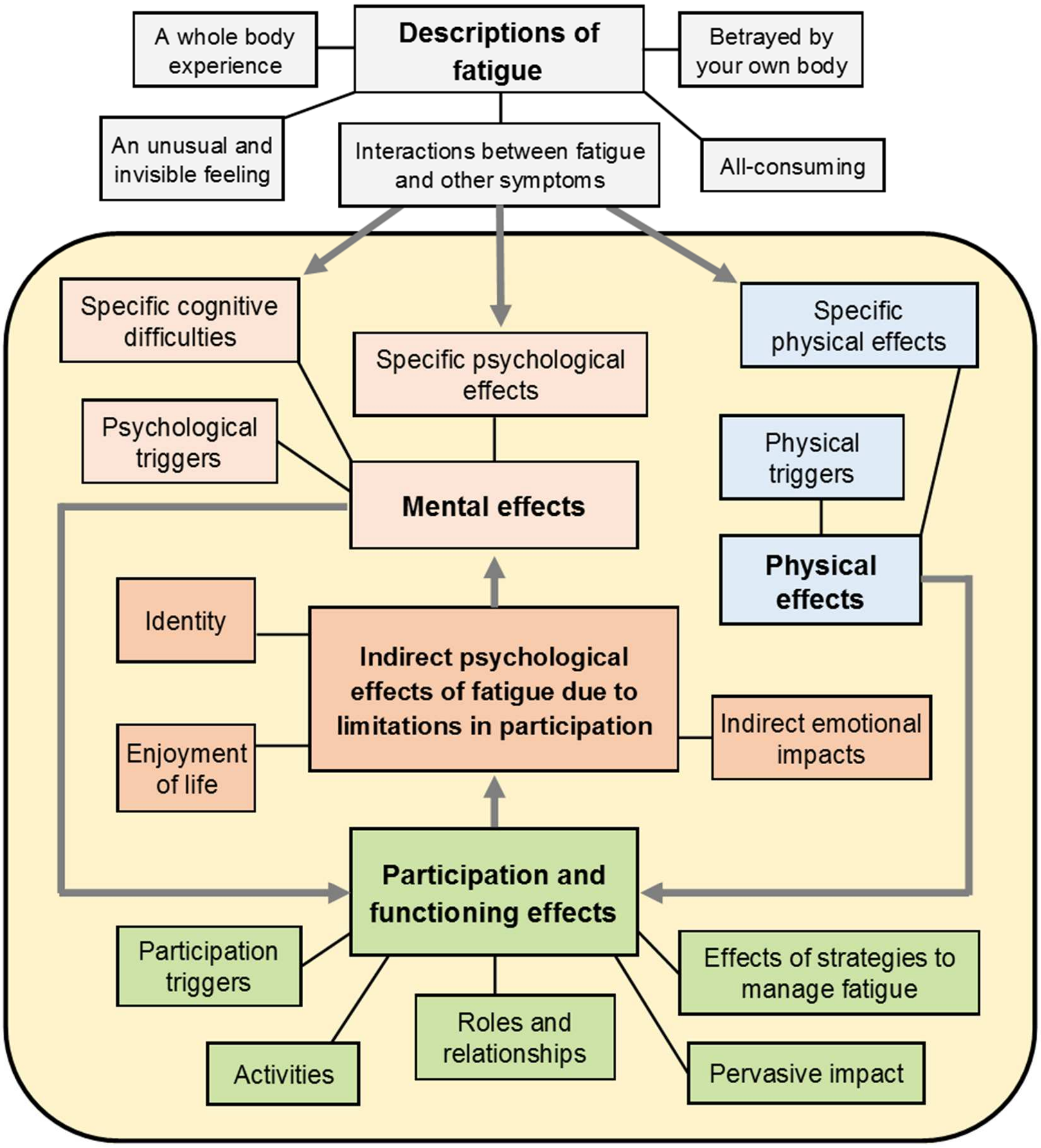


Figure A2: Conceptual framework of the impact of fatigue on people with MS

References

Cochrane 2017: http://msrdcns.cochrane.org/our-review, accessed 15/06/2017

Elbers RG, Verhoef J, van Wegen EEH, Berendse HW, Kwakkel G: Interventions for fatigue in Parkinson's disease. 2014: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010925/full accessed 22/07/2015

European Medicines Agency. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. EMEA/CHMP/EWP/139391/2004. London: European Medicines Agency; 2005.

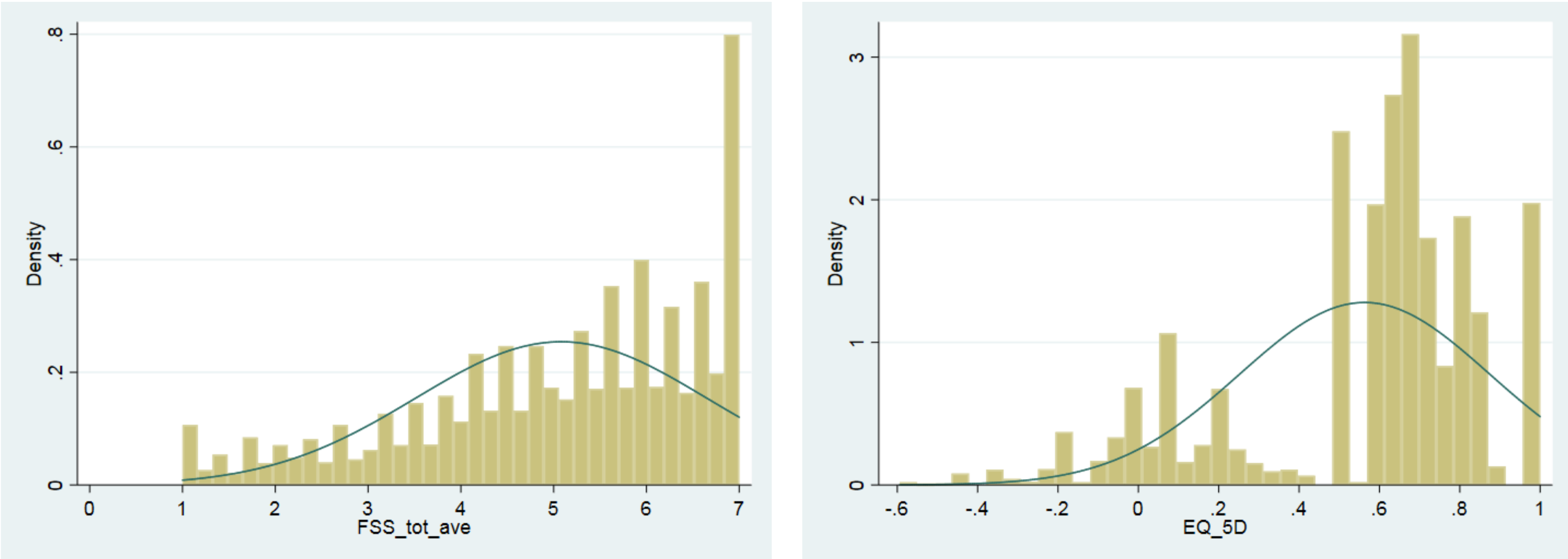
Riazi A. Patient-reported Outcome Measures in Multiple Sclerosis. The International MS Journal. 2006;13:92–99.

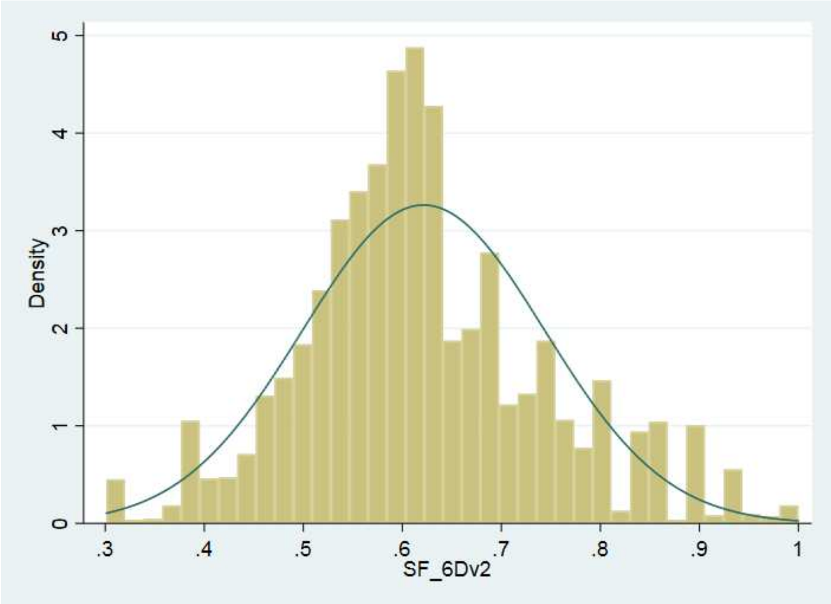
Shaw RL, Booth A, Sutton AJ, Miller T, Smith JA, Young B, Jones DR, Dixon-Woods M. Finding qualitative research: an evaluation of search strategies. BMC Medical Research Methodology 2004, 4:5

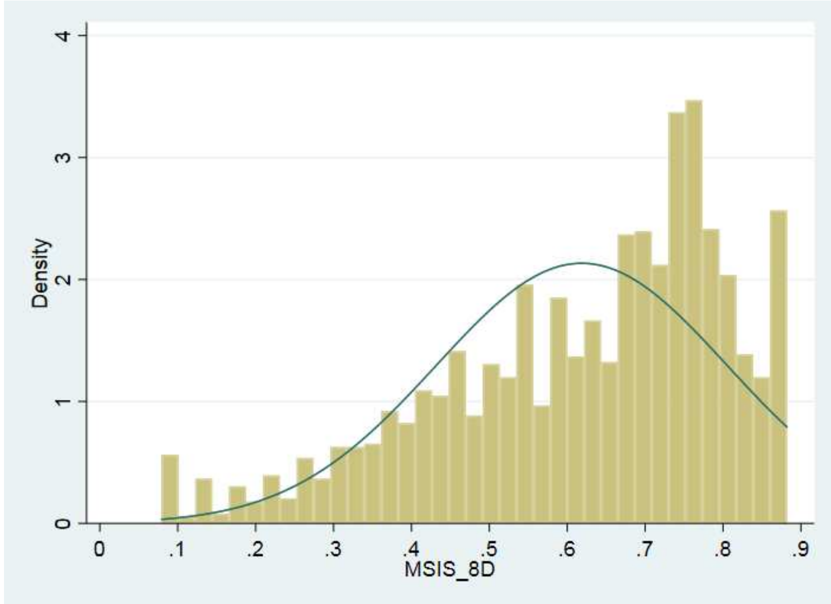
Ware J. Conceptualization and measurement of health-related quality of life: comments on an evolving field. Archives of Physical Medicine and Rehabilitation. 2003;84(Suppl 2):S43-S51.

Additional file 2

Histograms of source and target measures

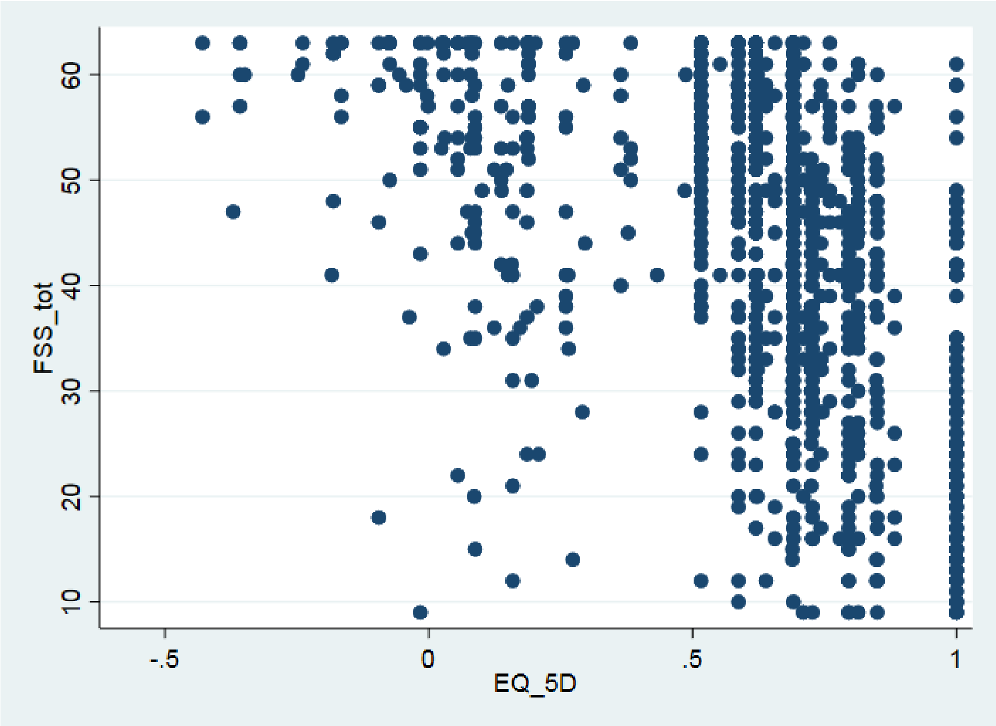



Density
SF_6Dv2

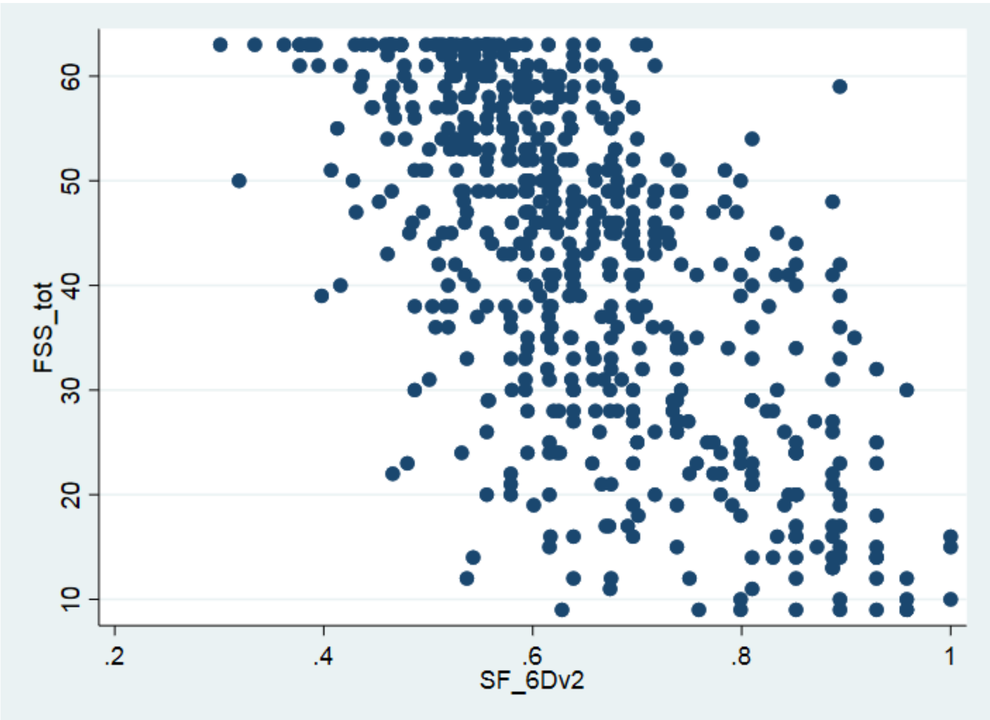

Density
MSIS_8D

Additional file 3

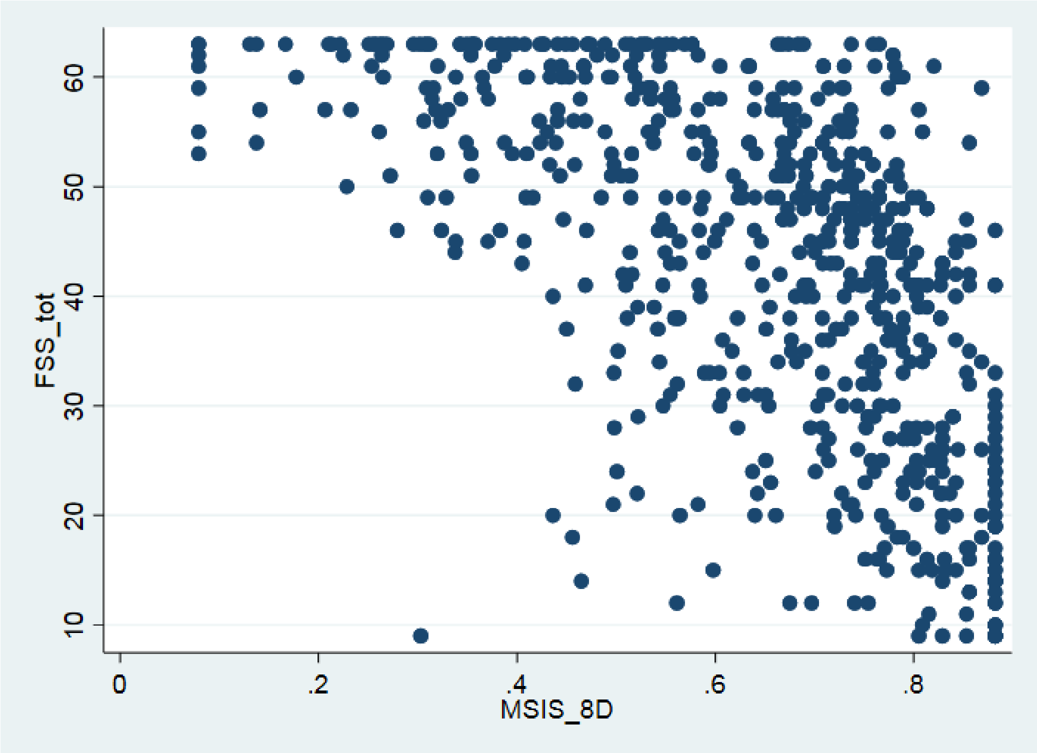
FSS total versus EQ-5D values



FSS total versus SF-6D values



FSS total versus MSIS-8D


FSS_tot
MSIS_8D
10
20
30
40
50
60
0
.2
.4
.6
.8

Additional file 4: All models run for this analysis

Initial models run using estimation dataset

OLS Model A: Regressing EQ-5D vs FSS total				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_tot	-0.008	0.001	-14.03	<0.0001
_cons	0.976	0.025	38.67	<0.0001
R2	0.2007			
RMSE	0.25087			
Coefficients	1			
Sig coefficients	1			
	Mean	Std.Dev.	Obs	
MAE	0.186	0.168	763	
MSE	0.063	0.113		
RMSE	0.251			
	Freq.	Percent		
Individuals with MAE < 0.25	577	75.62		
Individuals with MAE < 0.1	279	36.57		
Individuals with MAE < 0.05	150	19.66		
OLS Model B: Regressing EQ-5D vs FSS total and FSS total squared				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_tot	-0.002	0.003	-0.69	0.489
FSS_squared	0.000	0.000	-1.93	0.054
_cons	0.876	0.055	15.84	<0.0001
R2	0.2046			
RMSE	0.25043			
Coefficients	2			
Sig coefficients	0			
	Mean	Std.Dev.	Obs	
MAE	0.186	0.167	763	
MSE	0.062	0.112		
RMSE	0.250			
	Freq.	Percent		
Individuals with MAE < 0.25	578	75.75		
Individuals with MAE < 0.1	286	37.48		
Individuals with MAE < 0.05	155	20.31		
OLS Model C: Regressing EQ-5D vs FSS, age, gender				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_tot	-0.008	0.001	-13.86	<0.0001
age	-0.005	0.001	-6.39	<0.0001
female	-0.027	0.019	-1.37	0.170
_cons	1.235	0.043	28.43	<0.0001
R2	0.2465			
RMSE	0.24499			
Coefficients	3			

Sig coefficients	2			
	Mean	Std.Dev.	Obs	
MAE	0.184	0.161	755	
MSE	0.060	0.103		
RMSE	0.244			
	Freq.	Percent		
Individuals with MAE < 0.25	567	75.10		
Individuals with MAE < 0.1	281	37.22		
Individuals with MAE < 0.05	138	18.28		
OLS Model D: Regressing EQ-5D vs FSS item scores				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_01	0.012	0.007	1.84	0.066
FSS_02	-0.011	0.006	-1.87	0.062
FSS_03	-0.012	0.009	-1.32	0.186
FSS_04	-0.006	0.008	-0.68	0.499
FSS_05	-0.010	0.012	-0.84	0.401
FSS_06	-0.024	0.009	-2.73	0.007
FSS_07	-0.031	0.008	-3.96	<0.0001
FSS_08	0.031	0.006	4.73	<0.0001
FSS_09	-0.015	0.009	-1.75	0.081
_cons	0.901	0.032	28.44	<0.0001
R2	0.2584			
RMSE	0.24292			
Coefficients	9			
Sig coefficients	3			
	Mean	Std.Dev.	Obs	
MAE	0.182	0.159	763	
MSE	0.058	0.104		
RMSE	0.241			
	Freq.	Percent		
Individuals with MAE < 0.25	585	76.67		
Individuals with MAE < 0.1	287	37.61		
Individuals with MAE < 0.05	138	18.09		
OLS Model E: Regressing EQ-5D vs FSS item scores, age and gender				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_01	0.011	0.006	1.77	0.076
FSS_02	-0.011	0.006	-1.79	0.074
FSS_03	-0.014	0.009	-1.66	0.097
FSS_04	-0.005	0.008	-0.64	0.521
FSS_05	-0.010	0.011	-0.84	0.402
FSS_06	-0.020	0.009	-2.24	0.025
FSS_07	-0.028	0.008	-3.63	<0.0001
FSS_08	0.030	0.006	4.70	<0.0001
FSS_09	-0.017	0.008	-2.08	0.038

age	-0.004	0.001	-5.76	<0.0001
female	-0.039	0.019	-1.99	0.046
_cons	1.145	0.046	24.91	<0.0001
R2	0.2953			
RMSE	0.23819			
Coefficients	11			
Sig coefficients	6			
	Mean	Std.Dev.	Obs	
MAE	0.180	0.153	755	
MSE	0.056	0.097		
RMSE	0.236			
	Freq.	Percent		
Individuals with MAE < 0.25	585	77.48		
Individuals with MAE < 0.1	263	34.83		
Individuals with MAE < 0.05	137	18.15		
OLS Model A: Regressing SF-6D vs FSS total				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_tot	-0.006	0.000	-18.91	<0.0001
_cons	0.897	0.015	59.44	<0.0001
R2	0.4511			
RMSE	0.09985			
Coefficients	1			
Sig coefficients	1			
	Mean	Std.Dev.	Obs	
MAE	0.078	0.062	455	
MSE	0.010	0.015		
RMSE	0.100			
	Freq.	Percent		
Individuals with MAE < 0.25	449	98.68		
Individuals with MAE < 0.1	310	68.13		
Individuals with MAE < 0.05	188	41.32		
OLS Model B: Regressing SF-6D vs FSS total and FSS total squared				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_tot	-0.007	0.002	-4.31	<0.0001
FSS_squared	0.000	0.000	0.91	0.364
_cons	0.921	0.032	28.91	<0.0001
R2	0.4521			
RMSE	0.09987			
Coefficients	2			
Sig coefficients	1			
	Mean	Std.Dev.	Obs	
MAE	0.077	0.063	455	

MSE	0.010	0.015		
RMSE	0.100			
	Freq.	Percent		
Individuals with MAE < 0.25	449	98.68		
Individuals with MAE < 0.1	312	68.57		
Individuals with MAE < 0.05	193	42.42		
OLS Model C: Regressing SF-6D vs FSS, age, gender				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_tot	-0.006	0.000	-19.07	<0.0001
age	-0.001	0.000	-2.64	0.009
female	-0.026	0.010	-2.5	0.013
_cons	0.970	0.026	37.29	<0.0001
R2	0.4668			
RMSE	0.09888			
Coefficients	3			
Sig coefficients	3			
	Mean	Std.Dev.	Obs	
MAE	0.078	0.060	452	
MSE	0.010	0.014		
RMSE	0.098			
	Freq.	Percent		
Individuals with MAE < 0.25	445	98.45		
Individuals with MAE < 0.1	309	68.36		
Individuals with MAE < 0.05	181	40.04		
OLS Model D: Regressing SF-6D vs FSS item scores				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_01	-0.002	0.004	-0.55	0.580
FSS_02	-0.005	0.004	-1.38	0.168
FSS_03	-0.005	0.005	-0.99	0.323
FSS_04	0.001	0.005	0.22	0.830
FSS_05	-0.007	0.005	-1.5	0.135
FSS_06	-0.002	0.005	-0.5	0.618
FSS_07	-0.022	0.004	-4.94	<0.0001
FSS_08	0.008	0.004	2.14	0.033
FSS_09	-0.015	0.005	-3.06	0.002
_cons	0.871	0.018	49.25	<0.0001
R2	0.4872			
RMSE	0.09738			
Coefficients	9			
Sig coefficients	3			
	Mean	Std.Dev.	Obs	
MAE	0.075	0.061	455	
MSE	0.009	0.014		
RMSE	0.096			
	Freq.	Percent		

Individuals with MAE < 0.25	448	98.46		
Individuals with MAE < 0.1	326	71.65		
Individuals with MAE < 0.05	193	42.42		
OLS Model E: Regressing SF-6D vs FSS item scores, age and gender				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_01	-0.002	0.004	-0.48	0.630
FSS_02	-0.005	0.004	-1.46	0.145
FSS_03	-0.006	0.005	-1.07	0.283
FSS_04	0.001	0.005	0.13	0.898
FSS_05	-0.007	0.005	-1.43	0.153
FSS_06	-0.001	0.005	-0.25	0.805
FSS_07	-0.022	0.004	-5.04	<0.0001
FSS_08	0.009	0.004	2.33	0.020
FSS_09	-0.016	0.005	-3.43	0.001
age	-0.001	0.000	-2.39	0.017
female	-0.030	0.011	-2.81	0.005
_cons	0.943	0.029	32.57	<0.0001
R2	0.5043			
RMSE	0.0962			
Coefficients	11			
Sig coefficients	5			
	Mean	Std.Dev.	Obs	
MAE	0.074	0.059	452	
MSE	0.009	0.014		
RMSE	0.095			
	Freq.	Percent		
Individuals with MAE < 0.25	445	98.45		
Individuals with MAE < 0.1	326	72.12		
Individuals with MAE < 0.05	195	43.14		
OLS Model A: Regressing MSIS-8D vs FSS total				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_tot	-0.007	0.000	-17.29	<0.0001
_cons	0.961	0.016	58.69	<0.0001
R2	0.3665			
RMSE	0.15046			
Coefficients	1			
Sig coefficients	1			
	Mean	Std.Dev.	Obs	
MAE	0.119	0.092	493	
MSE	0.023	0.034		
RMSE	0.150			
	Freq.	Percent		
Individuals with MAE < 0.25	448	90.87		

Individuals with MAE < 0.1	242	49.09		
Individuals with MAE < 0.05	136	27.59		
OLS Model B: Regressing MSIS-8D vs FSS total and FSS total squared				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_tot	0.003	0.002	1.67	0.096
FSS_squared	0.000	0.000	-5.08	<0.0001
_cons	0.785	0.033	23.82	<0.0001
R2	0.3938			
RMSE	0.14734			
Coefficients	2			
Sig coefficients	1			
	Mean	Std.Dev.	Obs	
MAE	0.116	0.090	493	
MSE	0.022	0.033		
RMSE	0.147			
	Freq.	Percent		
Individuals with MAE < 0.25	447	90.67		
Individuals with MAE < 0.1	257	52.13		
Individuals with MAE < 0.05	118	23.94		
OLS Model C: Regressing MSIS-8D vs FSS, age, gender				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_tot	-0.007	0.000	-16.96	<0.0001
age	-0.001	0.001	-1.86	0.064
female	-0.008	0.016	-0.53	0.595
_cons	1.019	0.032	32.01	<0.0001
R2	0.3718			
RMSE	0.15048			
Coefficients	3			
Sig coefficients	1			
	Mean	Std.Dev.	Obs	
MAE	0.118	0.093	490	
MSE	0.022	0.034		
RMSE	0.150			
	Freq.	Percent		
Individuals with MAE < 0.25	447	91.22		
Individuals with MAE < 0.1	243	49.59		
Individuals with MAE < 0.05	133	27.14		
OLS Model D: Regressing MSIS-8D vs FSS item scores				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_01	0.001	0.005	0.12	0.902
FSS_02	-0.005	0.005	-1.02	0.307
FSS_03	-0.009	0.006	-1.45	0.148
FSS_04	0.006	0.006	1.07	0.287
FSS_05	-0.024	0.007	-3.29	0.001

FSS_06	-0.006	0.007	-0.81	0.417
FSS_07	-0.027	0.006	-4.5	<0.0001
FSS_08	0.025	0.004	5.81	<0.0001
FSS_09	-0.023	0.006	-3.82	<0.0001
_cons	0.902	0.020	44.71	<0.0001
R2	0.4389			
RMSE	0.14277			
Coefficients	9			
Sig coefficients	4			
	Mean	Std.Dev.	Obs	
MAE	0.109	0.089	493	
MSE	0.020	0.032		
RMSE	0.141			
	Freq.	Percent		
Individuals with MAE < 0.25	455	92.29		
Individuals with MAE < 0.1	265	53.75		
Individuals with MAE < 0.05	155	31.44		
OLS Model E: Regressing MSIS-8D vs FSS item scores, age and gender				
EQ-5D	Coefficient	SE	t	P>|t|
FSS_01	0.001	0.005	0.15	0.883
FSS_02	-0.005	0.005	-1.01	0.311
FSS_03	-0.009	0.006	-1.47	0.143
FSS_04	0.006	0.006	1.04	0.299
FSS_05	-0.024	0.007	-3.33	0.001
FSS_06	-0.005	0.007	-0.65	0.515
FSS_07	-0.027	0.006	-4.51	<0.0001
FSS_08	0.024	0.004	5.74	<0.0001
FSS_09	-0.023	0.006	-3.94	<0.0001
age	-0.001	0.001	-1.22	0.222
female	-0.014	0.016	-0.87	0.387
_cons	0.946	0.034	27.57	<0.0001
R2	0.4419			
RMSE	0.14301			
Coefficients	11			
Sig coefficients	4			
	Mean	Std.Dev.	Obs	
MAE	0.109	0.090	490	
MSE	0.020	0.033		
RMSE	0.141			
	Freq.	Percent		
Individuals with MAE < 0.25	455	92.86		
Individuals with MAE < 0.1	261	53.27		
Individuals with MAE < 0.05	157	32.04		

Note that FSS-08 (fatigue is among my most disabling symptoms) has a positive coefficient
FSS-01 (my motivation is lower when I am fatigued) has a positive coefficient in the EQ-5D and MSIS-8D models, but not in the SF-6D models

Models run using estimation dataset, included significant FSS items only

EQ-5D MODELS						
***CLAD Model C2**						
Variable	Observed	Bias	Std.Err.	LCL	UCL	
FSS total score	-0.00576	-9.1E-05	0.000608	-0.00696	-0.00455	
Age	-0.00309	-4.2E-05	0.000718	-0.00451	-0.00166	
Constant	1.084261	0.004113	0.049776	0.985494	1.183028	
Observations	755					
	Mean	Std.Dev.				
MAE	0.173114	0.183672				
MSE	0.063659	0.127679				
RMSE	0.252					
Pseudo R2	0.126022					
	Freq.	Percent				
Individuals with MAE < 0.25	597	79.07				
Individuals with MAE < 0.10	353	46.75				
Individuals with MAE < 0.05	190	25.17				
***CLAD Model D2**						
Variable	Observed	Bias	Std.Err.	LCL	UCL	
FSS_06	-0.026	0.002269	0.004559	-0.03505	-0.01695	
FSS_07	-0.02067	-0.00229	0.005582	-0.03174	-0.00959	
Constant	0.894667	-0.00099	0.016949	0.861036	0.928298	
Observations	774					
	Mean	Std.Dev.				
MAE	0.172271	0.186334				
MSE	0.064353	0.131224				
RMSE	0.254					
Pseudo R2	0.121661					
	Freq.	Percent				
Individuals with MAE < 0.25	608	78.55				
Individuals with MAE < 0.10	371	47.93				
Individuals with MAE < 0.05	221	28.55				
***CLAD Model E2**						
Variable	Observed	Bias	Std.Err.	LCL	UCL	
FSS_07	-0.02939	0.001426	0.00573	-0.04076	-0.01802	
FSS_09	-0.01681	-0.00099	0.004567	-0.02588	-0.00775	
Age	-0.00302	1.63E-05	0.000681	-0.00437	-0.00167	
Constant	1.038418	-0.00442	0.048066	0.943045	1.133792	

Observations	765					
	Mean	Std.Dev.				
MAE	0.171062	0.189995				
MSE	0.065313	0.134326				
RMSE	0.255565					
Pseudo R2	0.134896					
	Freq.	Percent				
Individuals with MAE < 0.25	607	79.35				
Individuals with MAE < 0.10	371	47.93				
Individuals with MAE < 0.05	222	29.02				
SF-6D MODELS						
***OLS Model D2**						
SF_6Dv2	Coef.	Std.Err.	t	P>|t|	LCL	UCL
FSS_07	-0.02902	0.003348	-8.67	<0.0001	-0.0356	-0.02244
FSS_09	-0.01699	0.00333	-5.1	<0.0001	-0.02353	-0.01045
Constant	0.852329	0.012796	66.61	<0.0001	0.827182	0.877476
Observations	460					
	Mean	Std.Dev.				
MAE	0.076699	0.061377				
MSE	0.009642	0.01459				
RMSE	0.098192					
R2	0.4639					
	Freq.	Percent				
Individuals with MAE < 0.25	454	98.7				
Individuals with MAE < 0.10	323	70.22				
Individuals with MAE < 0.05	193	41.96				
F-stat	180.27					
Prob	0					
***OLS Model E2**						
SF_6Dv2	Coef.	Std.Err.	t	P>|t|	LCL	UCL
FSS_07	-0.02864	0.003345	-8.56	<0.0001	-0.03521	-0.02206
FSS_09	-0.01747	0.003259	-5.36	<0.0001	-0.02387	-0.01106
Age	-0.00095	0.000418	-2.28	0.023	-0.00177	-0.00013
Gender (female)	-0.02583	0.010477	-2.47	0.014	-0.04642	-0.00524
Constant	0.919073	0.024879	36.94	<0.0001	0.87018	0.967966
Observations	457					
	Mean	Std.Dev.				
MAE	0.076046	0.060262				
MSE	0.009406	0.013951				
RMSE	0.096987					
R2	0.4796					
	Freq.	Percent				
Individuals with MAE < 0.25	448	98.03				

Individuals with MAE < 0.10	318	69.58				
Individuals with MAE < 0.05	184	40.26				
F-stat	98.23					
Prob	0					
***CLAD Model D2**						
Variable	Observed	Bias	Std.Err.	LCL	UCL	
FSS_07	-0.03	-0.00061	0.004572	-0.03907	-0.02093	
FSS_09	-0.018	0.000174	0.004707	-0.02734	-0.00866	
Constant	0.858	0.003514	0.018524	0.821245	0.894755	
Observations	460					
	Mean	Std.Dev.				
MAE	0.076571	0.061931				
MSE	0.00969	0.01486				
RMSE	0.098					
Pseudo R2	0.275523					
	Freq.	Percent				
Individuals with MAE < 0.25	452	98.26				
Individuals with MAE < 0.10	325	70.65				
Individuals with MAE < 0.05	190	41.3				
***CLAD Model E2**						
Variable	Observed	Bias	Std.Err.	LCL	UCL	
FSS_07	-0.03085	0.000214	0.004454	-0.03969	-0.02201	
FSS_09	-0.0179	-0.00062	0.004468	-0.02677	-0.00903	
Gender (female)	-0.0215	-0.00175	0.013233	-0.04776	0.004757	
Constant	0.90175	0.004941	0.031155	0.839932	0.963568	
Observations	460					
	Mean	Std.Dev.				
MAE	0.076436	0.061926				
MSE	0.009669	0.014696				
RMSE	0.09833					
Pseudo R2	0.279408					
	Freq.	Percent				
Individuals with MAE < 0.25	453	98.48				
Individuals with MAE < 0.10	324	70.43				
Individuals with MAE < 0.05	201	43.7				
MSIS-8D MODELS						
***OLS Model D2**						
MSIS_8D	Coef.	Std.Err.	t	P>|t|	LCL	UCL
FSS_05	-0.02755	0.005173	-5.33	<0.0001	-0.03771	-0.01738
FSS_07	-0.03342	0.00496	-6.74	<0.0001	-0.04317	-0.02368
Constant	0.914556	0.012816	71.36	<0.0001	0.889375	0.939736

Observations	500					
	Mean	Std.Dev.				
MAE	0.113603	0.090267				
MSE	0.021038	0.032955				
RMSE	0.145043					
R2	0.4048					
	Freq.	Percent				
Individuals with MAE < 0.25	459	91.8				
Individuals with MAE < 0.10	262	52.4				
Individuals with MAE < 0.05	138	27.6				
F-stat	170.87					
Prob	0					
*CLAD Model D2						
Variable	Observed	Bias	Std.Err.	LCL	UCL	
FSS_07	-0.05296	-0.00163	0.003466	-0.05983	-0.04608	
Constant	0.908928	0.00911	0.013695	0.881754	0.936102	
Observations	501					
	Mean	Std.Dev.				
MAE	0.114843	0.102382				
MSE	0.02365	0.040465				
RMSE	0.154					
Pseudo R2	0.223803					
	Freq.	Percent				
Individuals with MAE < 0.25	449	89.62				
Individuals with MAE < 0.10	271	54.09				
Individuals with MAE < 0.05	154	30.74				
CLAD MODELS WITH P VALUES						
CLAD_eq_C2.smcl						
	Coef.	Std. Err	z	P>|z|	LCL	UCL
FSS_tot	-0.00576	0.000608	-9.47	<0.0001	-0.00695	-0.00456
age	-0.00309	0.000718	-4.3	<0.0001	-0.00449	-0.00168
const	1.084261	0.049776	21.78	<0.0001	0.986701	1.181821
CLAD_eq_D2.smcl						
	Coef.	Std. Err	z	P>|z|	LCL	UCL
FSS_06	-0.026	0.004559	-5.7	<0.0001	-0.03493	-0.01707
FSS_07	-0.02067	0.005582	-3.7	<0.0001	-0.03161	-0.00973
const	0.894667	0.016949	52.79	<0.0001	0.861447	0.927886
CLAD_eq_E2.smcl						
	Coef.	Std. Err	z	P>|z|	LCL	UCL

FSS_07	-0.02939	0.00573	-5.13	<0.0001	-0.04062	-0.01816
FSS_09	-0.01681	0.004567	-3.68	<0.0001	-0.02577	-0.00786
age	-0.00302	0.000681	-4.43	<0.0001	-0.00435	-0.00168
const	1.038418	0.048066	21.6	<0.0001	0.94421	1.132626
CLAD_sf_D2.smcl						
	Coef.	Std. Err	z	P>|z|	LCL	UCL
FSS_07	-0.03	0.004572	-6.56	<0.0001	-0.03896	-0.02104
FSS_09	-0.018	0.004707	-3.82	<0.0001	-0.02723	-0.00878
const	0.858	0.018524	46.32	<0.0001	0.821694	0.894306
CLAD_sf_E2.smcl						
	Coef.	Std. Err	z	P>|z|	LCL	UCL
FSS_07	-0.03085	0.004454	-6.93	<0.0001	-0.03958	-0.02212
FSS_09	-0.0179	0.004468	-4.01	<0.0001	-0.02666	-0.00914
Gender	-0.0215	0.013233	-1.62	0.104	-0.04744	0.004436
const	0.90175	0.031155	28.94	<0.0001	0.840688	0.962812
CLAD_ms_D2.smcl						
	Coef.	Std. Err	z	P>|z|	LCL	UCL
FSS_07	-0.05296	0.003466	-15.28	<0.0001	-0.05975	-0.04616
const	0.908928	0.013695	66.37	<0.0001	0.882086	0.93577

All models run using validation dataset

EQ-5D CLAD MODEL A							
Variable	Observed	Bias	SE	LCL	UCL	z	P>|z|
FSS total score	-0.00726	-0.00018	0.001184	-0.00966	-0.00528	-6.14	<0.0001
Constant	1.001	0.005987	0.054856	0.9133	1.0968	18.25	<0.0001
Observations	260						
	Mean	SD					
MAE	0.183491	0.206009					
MSE	0.075946	0.165228					
RMSE	0.276						
Pseudo R2	0.119159						
	Freq.	Percent					
Individuals with MAE<0.25	205	78.85					
Individuals with MAE<0.10	129	49.62					
Individuals with MAE<0.05	64	24.62					
EQ-5D CLAD MODEL C							

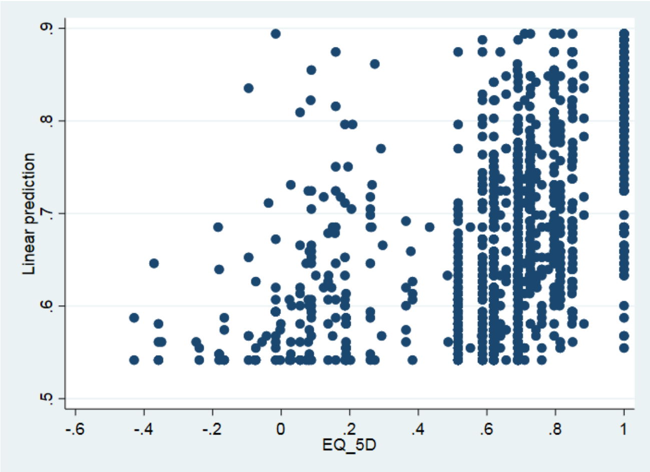
Variable	Observed	Bias	SE	LCL	UCL	z	P>|z|
FSS total score	-0.00757	0.000273	0.001066	-0.01062	-0.00559	-7.1	<0.0001
Age	-0.00423	0.000525	0.001077	-0.0056	-0.00245	-3.92	<0.0001
Gender (female)	-0.00898	0.017491	0.025959	-0.04416	0.033484	-0.35	0.729
Constant	1.233316	-0.06679	0.09793	1.094404	1.353203	12.59	<0.0001
Observations	260						
	Mean	SD					
MAE	0.178599	0.198645					
MSE	0.071206	0.153685					
RMSE	0.267						
Pseudo R2	0.141096						
	Freq.	Percent					
Individuals with MAE<0.25	200	76.92					
Individuals with MAE<0.10	123	47.31					
Individuals with MAE<0.05	67	25.77					
SF-6D OLS MODEL A							
SF_6Dv2	Coef.	SE	t	P>t	LCL	UCL	
FSS total score	-0.00401	0.000542	-7.4	0	-0.00508	-0.00294	
Constant	0.809539	0.026061	31.06	0	0.758044	0.861033	
Observations	152						
	Mean	SD					
MAE	0.068365	0.058008					
MSE	0.008017	0.012791					
RMSE	0.089535						
R2	0.3155						
	Freq.	Percent					
Individuals with MAE<0.25	150	98.68					
Individuals with MAE<0.10	116	76.32					
Individuals with MAE<0.05	74	48.68					
F	54.71						
Prob	0						
SF-6D CLAD MODEL A							
Variable	Observed	Bias	SE	LCL	UCL	z	P>|z|
FSS total score	-0.00377	-0.00017	0.000729	-0.00519	-0.002	-5.18	<0.0001
Constant	0.792955	0.009771	0.039386	0.738167	0.886	20.13	<0.0001
Observations	152						
	Mean	SD					
MAE	0.068255	0.059064					
MSE	0.008124	0.012945					
RMSE	0.09						

Pseudo R2	0.168562						
	Freq.	Percent					
Individuals with MAE<0.25	149	98.03					
Individuals with MAE<0.10	114	75					
Individuals with MAE<0.05	70	46.05					
SF-6D CLAD MODEL C							
Variable	Observed	Bias	SE	LCL	UCL	z	P>|z|
FSS total score	-0.0041	0.000174	0.000789	-0.00546	-0.00302	-5.19	<0.0001
Age	-0.00043	7.25E-05	0.000917	-0.00195	0.001836	-0.47	0.636
Gender (female)	0.00184	-0.00312	0.018711	-0.03795	0.035211	0.1	0.922
Constant	0.826868	-0.0029	0.078108	0.676031	0.992854	10.59	<0.0001
Observations	152						
	Mean	SD					
MAE	0.070781	0.063761					
MSE	0.009049	0.014786					
RMSE	0.095						
Pseudo R2	0.169288						
	Freq.	Percent					
Individuals with MAE<0.25	149	98.03					
Individuals with MAE<0.10	115	75.66					
Individuals with MAE<0.05	70	46.05					
MSIS-8D CLAD MODEL A							
Variable	Observed	Bias	SE	LCL	UCL	z	P>|z|
FSS total score	-0.00639	-2.8E-06	0.000966	-0.01004	-0.00519	-6.61	<0.0001
Constant	0.939022	0.005128	0.043165	0.903513	1.131484	21.75	<0.0001
Observations	157						
	Mean	SD					
MAE	0.118181	0.094082					
MSE	0.022762	0.044508					
RMSE	0.151						
Pseudo R2	0.180284						
	Freq.	Percent					
Individuals with MAE<0.25	145	92.36					
Individuals with MAE<0.10	79	50.32					
Individuals with MAE<0.05	36	22.93					
MSIS-8D CLAD MODEL C							

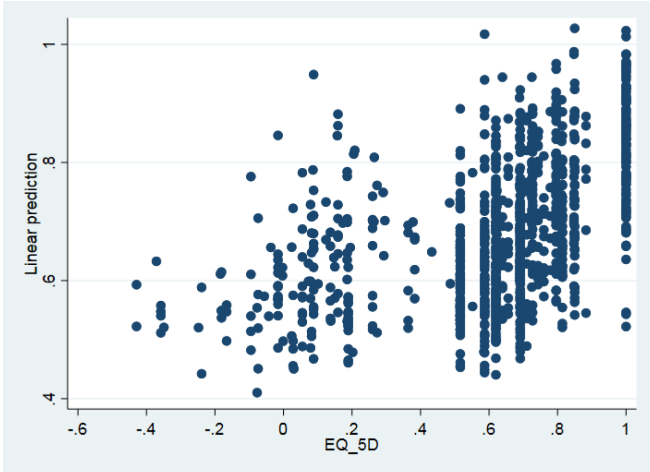
Variable	Observed	Bias	SE	LCL	UCL	z	P>|z|
FSS total score	-0.00597	-0.00019	0.001074	-0.00897	-0.00435	-5.56	<0.0001
Age	-0.00149	0.000546	0.002014	-0.00385	0.002588	-0.74	0.459
Gender (female)	0.011789	0.01073	0.039513	-0.0528	0.08738	0.3	0.765
Constant	0.974376	-0.02785	0.125225	0.728175	1.13045	7.78	<0.0001
Observations	157						
	Mean	SD					
MAE	0.114289	0.096382					
MSE	0.022292	0.035955					
RMSE	0.149						
Pseudo R2	0.185381						
	Freq.	Percent					
Individuals with MAE<0.25	143	91.08					
Individuals with MAE<0.10	82	52.23					
Individuals with MAE<0.05	49	31.21					

Additional file 5

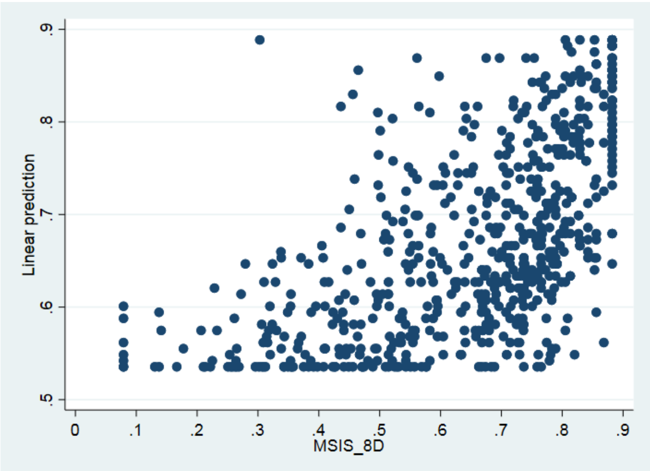
Scatterplots of observed vs predicted HSVs



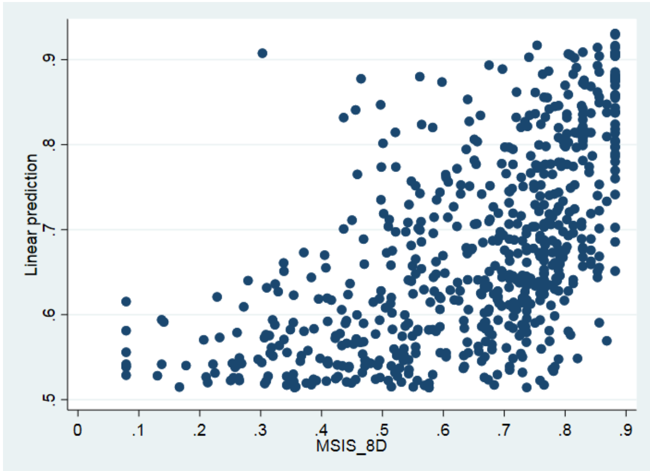
Observed EQ-5D vs EQ-5D estimated using CLAD A



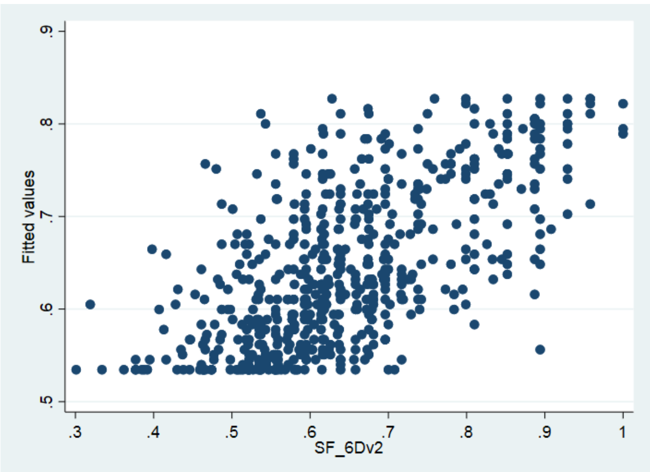
Observed EQ-5D vs EQ-5D estimated using CLAD C



Observed MSIS-8D vs MSIS-8D estimated using CLAD A

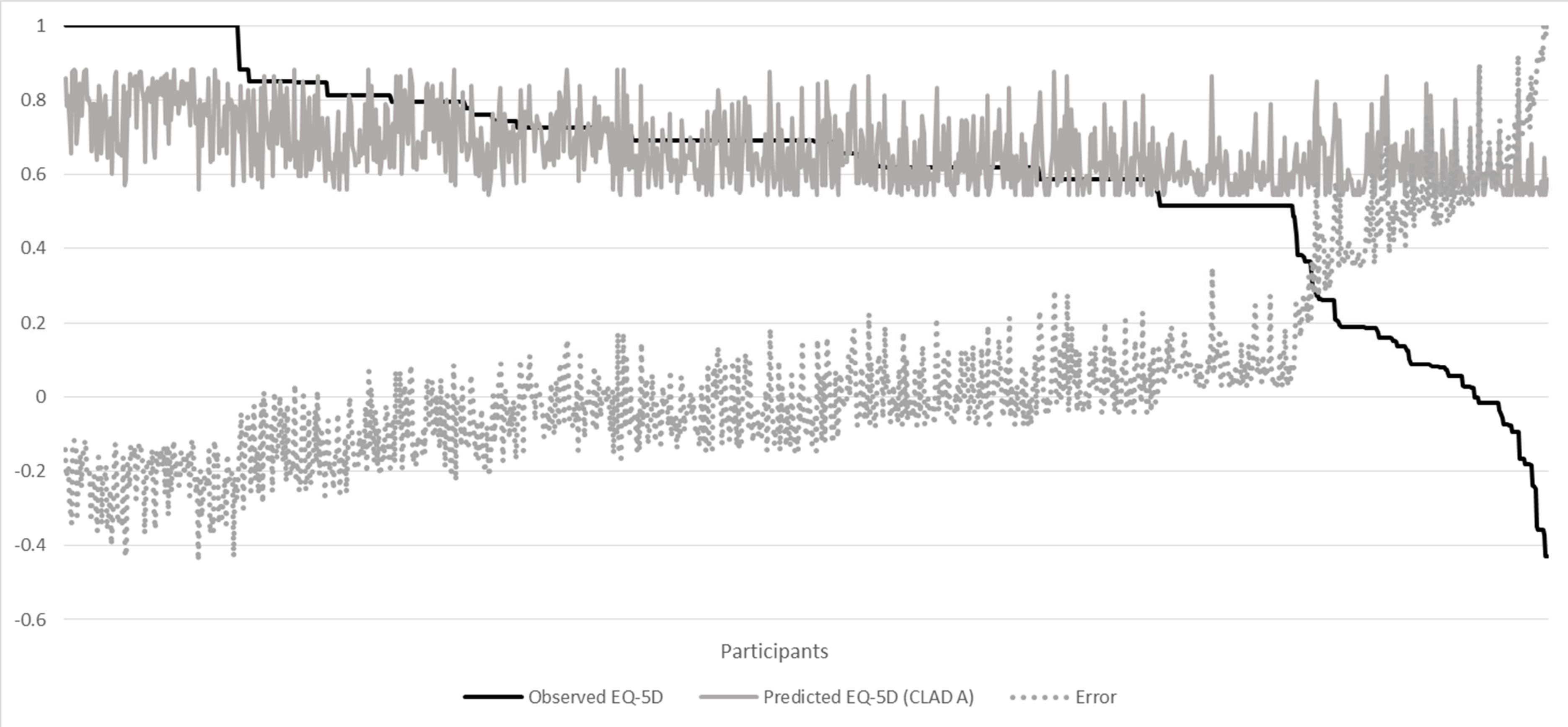


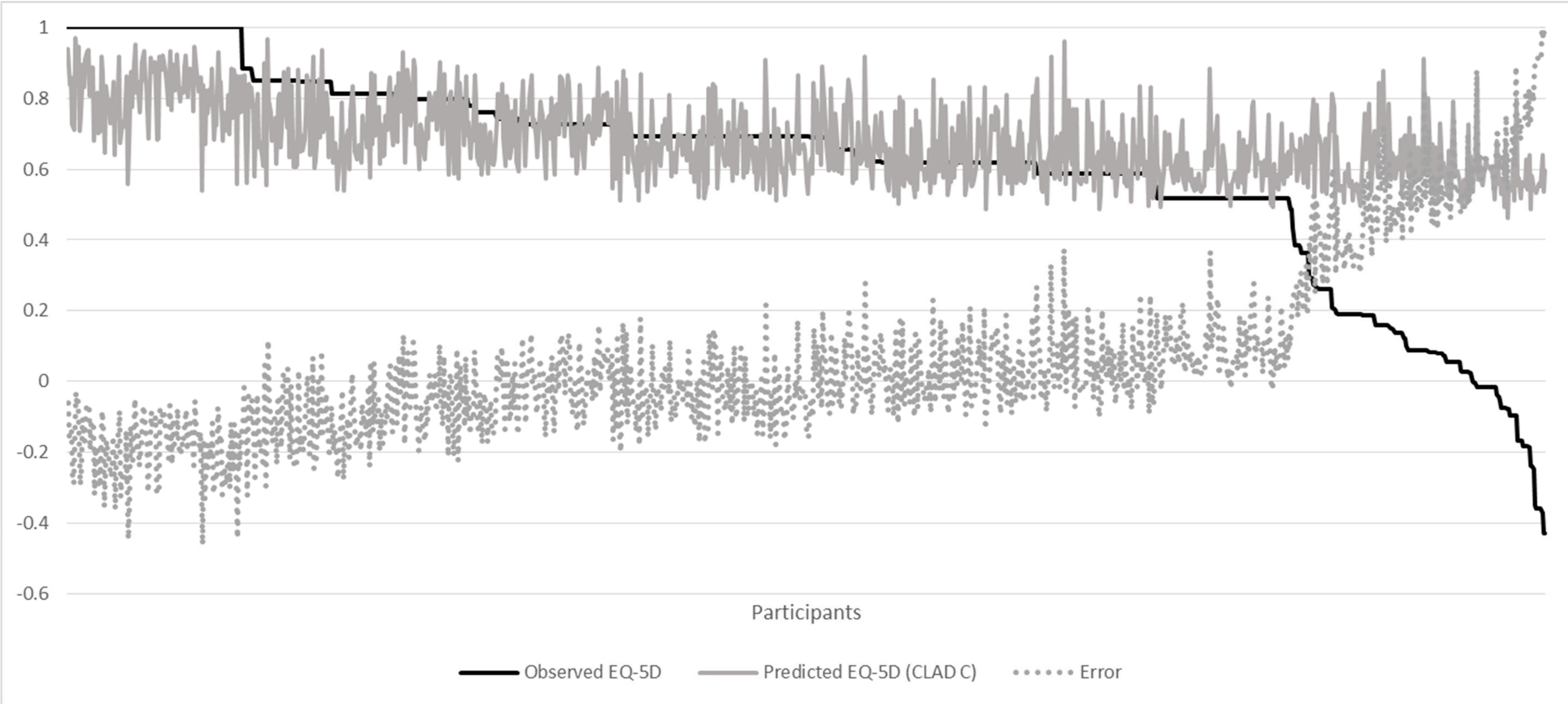
Observed MSIS-8D vs MSIS-8D estimated using CLAD C

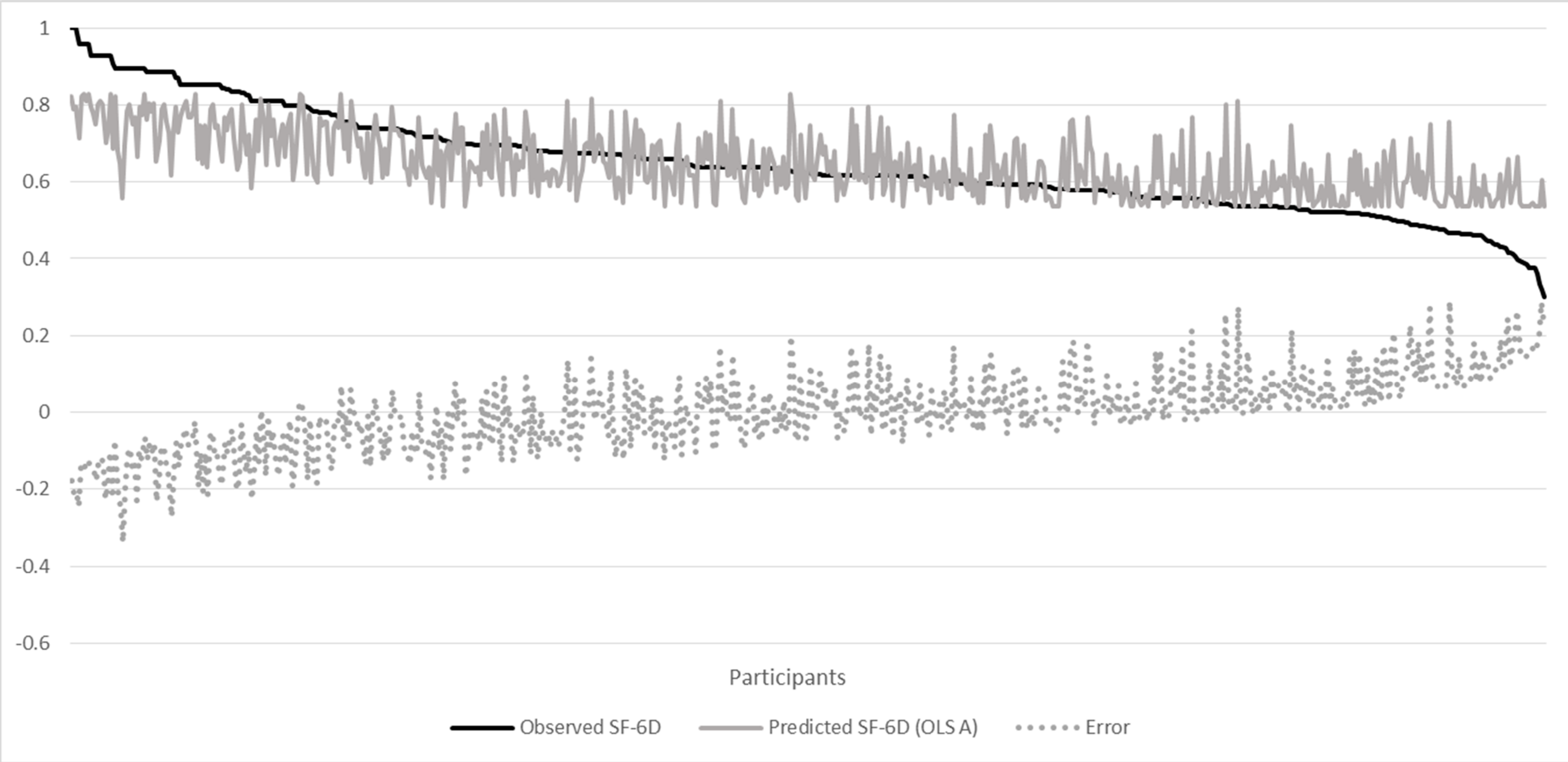


Observed SF-6D vs SF-6D estimated using OLS A

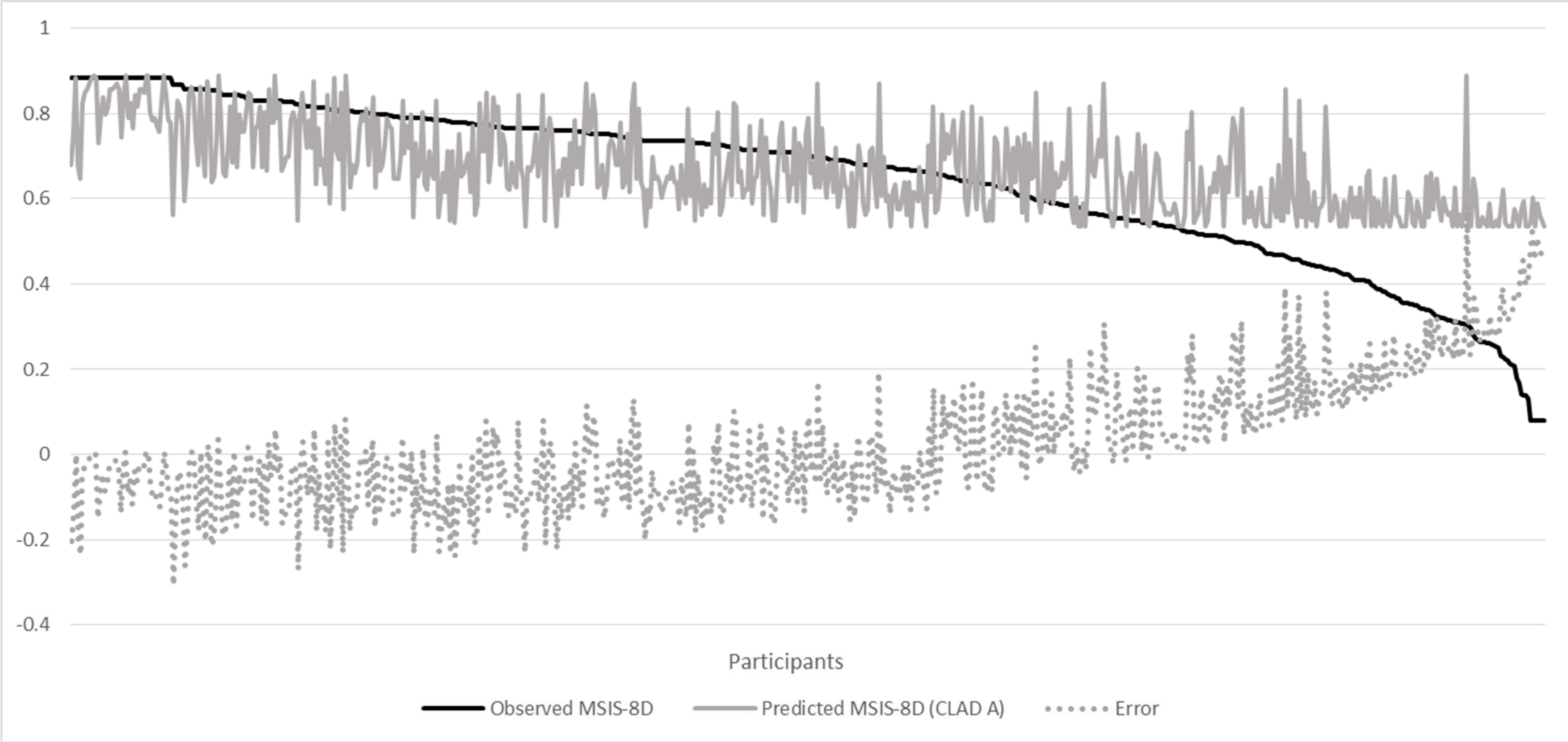
Additional file 6: Observed versus predicted HSVs

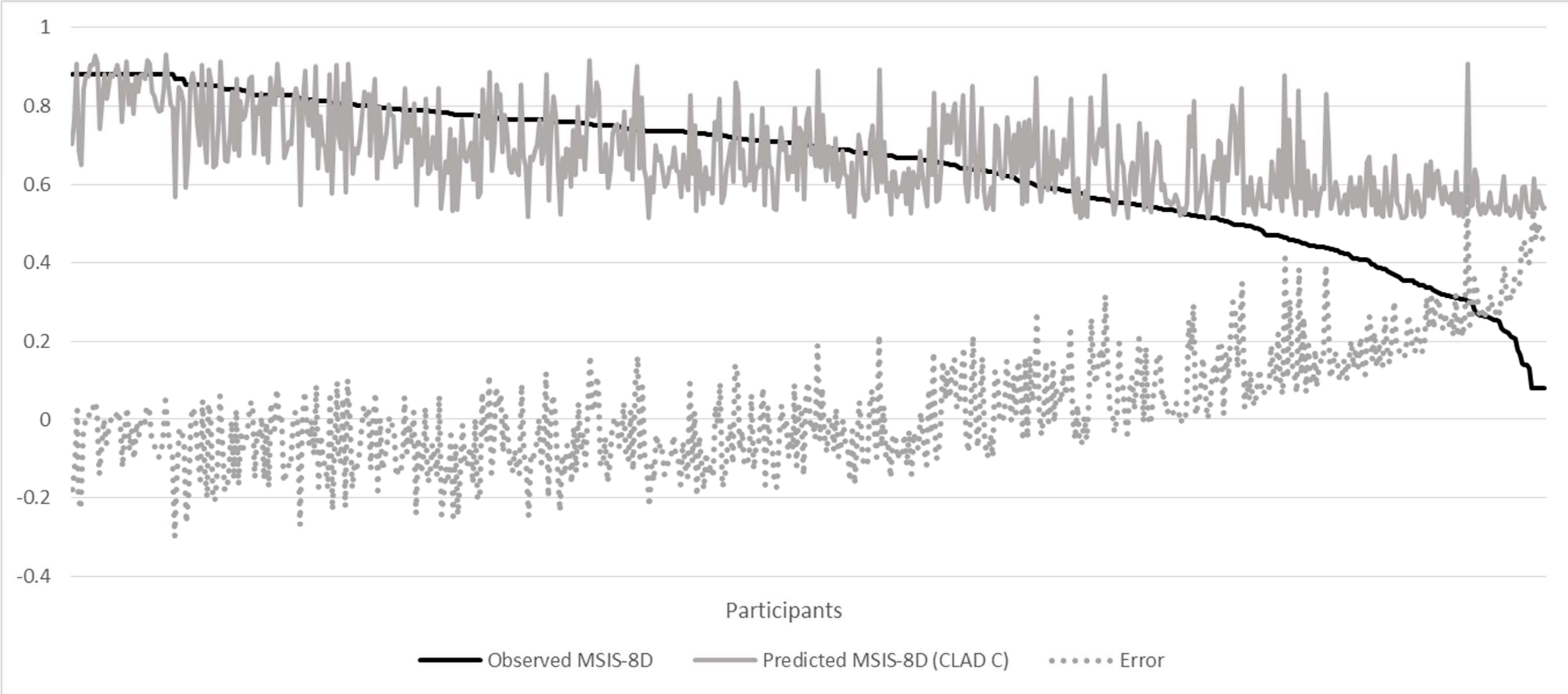



1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Participants
Observed EQ-5D
Predicted EQ-5D (CLAD C)
Error



1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Participants
Observed SF-6D
Predicted SF-6D (OLS A)
Error


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Participants
Observed MSIS-8D
Predicted MSIS-8D (CLAD A)
Error


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Participants
Observed MSIS-8D
Predicted MSIS-8D (CLAD C)
Error