

Impact on the Incremental Cost-Effectiveness Ratio of Using Alternatives to EQ-5D in a Markov Model for Multiple Sclerosis

Matthijs Versteegh¹

Published online: 9 June 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Objectives This study responds to a request in the National Institute for Health and Care Excellence (NICE) guidance to assess the impact of using alternative sources of utility values, applied to multiple sclerosis (MS).

Methods Incremental cost-effectiveness ratios (ICERs) were calculated using utility values based on UK and Dutch values of EQ-5D, two UK mappings and one Dutch mapping of EQ-5D and two condition-specific instruments: the UK eight-dimensional Multiple Sclerosis Impact Scale (MSIS-8D) and the Dutch Multiple Sclerosis Impact Scale Preference-Based Measure (MSIS-PBM). Deterministic and Monte-Carlo simulation-based ICERs were estimated for glatiramer acetate versus symptom management using a lifetime Markov model.

Results For both UK and Dutch perspectives, mapped and condition-specific utility values expressed significantly higher quality of life for the worst health state of the model than did EQ-5D. The ICER of glatiramer acetate with EQ-5D was US\$182,291 for The Netherlands and US\$153,476 for the UK. Ratios for mapped and condition-specific utilities were between 20 and 60 % higher.

Conclusion The overestimation of quality of life of patients with MS by mapped EQ-5D or condition-specific utility values, relative to observed EQ-5D, increases the ICER substantially in a lifetime Markov model.

Electronic supplementary material The online version of this article (doi:10.1007/s40273-016-0421-0) contains supplementary material, which is available to authorized users.

✉ Matthijs Versteegh
versteegh@imta.eur.nl

¹ Institute for Medical Technology Assessment, Erasmus University of Rotterdam, Burgemeester Oudlaan 50, 3000 DR Rotterdam, The Netherlands

Key Points for Decision Makers

The currently available mappings and condition-specific preference-based instruments can result in incremental cost-effectiveness ratios at least 20 % higher than those based on EQ-5D in a lifetime model for multiple sclerosis (MS).

Health technology assessment bodies should be aware that the different instruments to measure utilities in MS can result in rather different outcomes when applied in cost-utility analyses.

1 Introduction

Multiple sclerosis (MS) is a severely debilitating neurological condition affecting approximately 126,669 people in the UK, with about 6000 new cases each year [1]. Improvement of quality of life is a key treatment goal for this progressive condition, with limited impact on mortality. New pharmacological treatments that can improve the quality of life of patients with MS have to demonstrate a favourable outcome in economic evaluations before being eligible for reimbursement in both the UK and The Netherlands. In these economic evaluations, the health benefit of new pharmacological treatments is assessed with the quality-adjusted life-year (QALY). Computing health benefit with QALYs requires information on length of life and health-related quality of life (HR-QOL) in the form of utility values. HR-QOL can be measured with several instruments, of which EQ-5D is one of the most widely used. This study responds to a request in the National

Institute for Health and Care Excellence (NICE) guidance (paragraph 5.3.10 in the guidance [2] and section 2.4 of *Technical Support Document 11* [3]) to assess the impact on incremental cost-effectiveness ratios (ICERs) of using alternative sources of utility values (observed EQ-5D, mapped EQ-5D and condition-specific instruments) for economic evaluations in MS.

In the UK, NICE prefers EQ-5D-based utilities as a measure of HR-QOL. According to the NICE guidance, “when EQ-5D data are not available, these data can be estimated by mapping (...)” (paragraph 5.3.9) [2]. Also, when EQ-5D is not considered the most appropriate measure of HR-QOL, and when this consideration is substantiated with empirical data, alternatives to EQ-5D can be accepted, which is further described in *Technical Support Document 11* [3]. For both alternatives to directly observed EQ-5D data, the guidance calls for an assessment of using the alternative source of utility on the outcomes of the economic evaluation. Such an assessment is often complicated as EQ-5D is absent or less appropriate by definition when the alternatives are applied. The new Dutch guidelines for economic evaluations put an ever greater emphasis on EQ-5D than the UK guidelines. Alternatives to EQ-5D may be accepted, but even when EQ-5D is demonstrably not appropriate, alternative sources of utility values have to be presented alongside a base-case analysis that includes EQ-5D [4].

The preference of reimbursement authorities for one single generic preference-based measure as source of utility values has benefits in terms of achieving a uniform and comparable assessment of benefit. In case EQ-5D is absent, mapping to EQ-5D is preferred as this replicates EQ-5D data and contributes to this uniform assessment. However, the advantage of a uniform assessment comes at the potential cost of having a measure that is targeted at those elements of quality of life that are not targeted or “relevant and sensitive to those things that matter to patients with the condition” [5]. For example, fatigue and problems with cognition are not directly included in EQ-5D [6], while both are important symptoms for MS patients [7]. Therefore, two MS-specific preference-based measures developed in the past 5 years include an item on “feeling mentally fatigued” [6, 8]. It is uncertain, however, if alternatives to EQ-5D might be too focused on MS and may therefore miss important adverse effects or co-morbidities [5]. In other words, MS-specific instruments may be ‘too specific’ for an adequate assessment of the benefit of treatments in terms of HR-QOL.

To assess if a new condition-specific instrument has merit, the psychometric properties are compared with existing generic instruments such as EQ-5D to show increased sensitivity. What receives less attention, however, is that the ultimate goal of these instruments is to calculate QALYs for economic evaluations. As a

consequence, it is often overlooked that improvements in psychometric properties (i.e. favourable statistical properties such as larger effect sizes) might not be very relevant in the context of health economic modelling.

The evidence regarding the performance of the utility instruments available for MS can be summarised as follows: EQ-5D has limited discriminant validity for the middle Expanded Disability Status Scale (EDSS) categories, while mapped and condition-specific utilities have a floor effect compared with EQ-5D for patients in poor health. The three-level version of EQ-5D has limited discriminatory power between different categories of the EDSS, such as categories 3 and 4 [9] and categories 3–5, while it has been shown that the physical scale of the 29-item Multiple Sclerosis Impact Scale (MSIS-29) was able to discriminate between these categories [10]. These observations are confirmed in a large multi-country study [11]. The first alternative to EQ-5D is mapping, which only applies when EQ-5D is considered appropriate, something that is uncertain in MS. For MS, mapping algorithms have been developed and tested for the prediction of Dutch and UK utilities; these general show good performance but over-predict the utility values of patients in poor health [10, 12]. The second alternative to EQ-5D is a condition-specific instrument, two of which exist for MS: the Dutch Multiple Sclerosis Impact Scale Preference-Based Measure (MSIS-PBM) [8] and the UK eight-dimensional Multiple Sclerosis Impact Scale (MSIS-8D) [6]. The MSIS-PBM showed increased sensitivity at the cost of a limited capability to describe patients in poor health compared with EQ-5D [8]. Also, the lowest score on the MSIS-PBM was 0.42, which is high compared with the lowest score of the EQ-5D which was -0.329 (using the Dutch tariff) [13]. The MSIS-8D has a lowest attainable utility value of 0.08 but the instrument displays a ceiling effect compared with EQ-5D, as the best attainable value is 0.88 versus 1 with EQ-5D for patients in full health.

Based on the available evidence, all instruments that are currently available to calculate QALYs in MS have different properties and associated measurement issues. This study investigates how the properties of each of the instruments impact on the ICER of an economic evaluation in MS.

2 Methods

Economic evaluations of new treatment strategies for MS often use Markov models with health states based on EDSS categories [14–16]. For this study, the model structure, costs, effects and transition probabilities were taken from Bell et al. [16]. The model compares symptom

management with subcutaneous glatiramer acetate in MS patients with relapsing-remitting MS. First, utility values were calculated using data from the UK Multiple Sclerosis Risk-Sharing Scheme Monitoring Study, based on EQ-5D (Dutch and UK tariff), a mapping to Dutch EQ-5D utility values and two mappings to UK EQ-5D utility values, and the Dutch MSIS-PBM and UK MSIS-8D. Then, these utility values were applied to the health states of the Markov model. Seven ICERs were calculated, one for each of the instruments for Dutch and UK utilities.

2.1 Data Sources

This study combines a published Markov model with cross-sectional utility data ($n = 1295$) calculated using data from the UK risk-sharing monitoring scheme [17] using patient responses to EQ-5D and MSIS-29 questionnaires (detailed in Sects. 2.3.1 and 2.3.2). The cross-sectional data were collected from MS patients in 70 specialist centres in the UK.

2.2 Markov Model

The Markov model was based on a previously published model [16]. The model has a 1-month cycle length, a lifetime time horizon and seven health states: four EDSS-based health states, referred to as EDSS1 (EDSS score 0–2.5), EDSS2 (EDSS score 3–5.5), EDSS3 (EDSS score 6–7.5) and EDSS4 (EDSS score 8–9.5), two relapse health states differing in severity, and death. The model structure is depicted in Fig. 1. Medical management is compared to subcutaneous glatiramer acetate, which is only prescribed to patients in the EDSS1 and EDSS2 health states or for relapses experienced during those health states. The treatment effect was expected to reduce the probability of relapses as well as reduce the progression of the disease. The ICERs produced here should be interpreted in relative terms not absolute terms, as not all model aspects could be derived from the original publication. A description of the model and its parameters is provided in the Electronic Supplementary Material.

2.3 Sources of Utility Values for Quality-Adjusted Life-Year Calculation

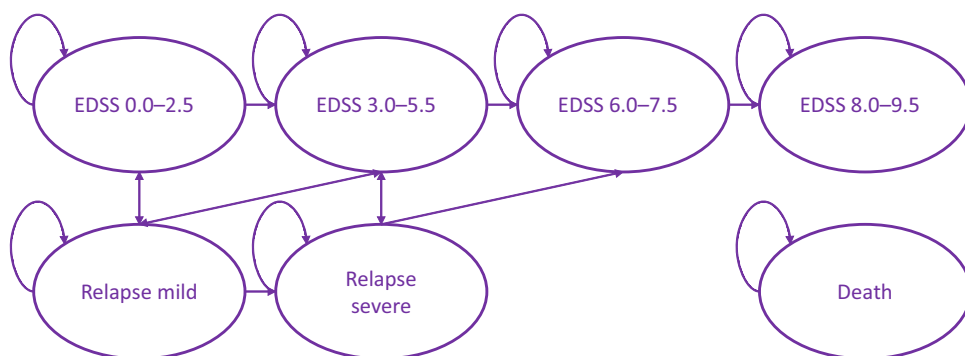
The calculation of QALYs requires an assessment of length of life and HR-QOL. Here, six sources of values for HR-QOL are described: the EQ-5D with UK and Dutch values, mapped EQ-5D for UK and Dutch values, and two condition-specific preference-based instruments, one for the UK and one for The Netherlands. All instruments were applied in the same risk-sharing scheme monitoring study database.

Utility values were calculated by applying the utility algorithms described in the following sections to the patient data of the UK risk-sharing monitoring scheme. Only patients that had data for all utility instruments were included. The mean and standard error (SE) for the four EDSS-based health states were then derived by clustering patients in four EDSS-based groups following the EDSS coding of the Markov model described in Sect. 2.2. The derived values were then applied to the Markov model to calculate QALYs. For the two relapse states, the same procedure as applied in the original publication of the Markov model was used; each relapse state was assigned a utility that was 0.094 lower than the mean utility of the state from which the patient entered the relapse state (i.e. EDSS1 or EDSS2). Quality of life was assumed equal for patients receiving symptom management and patients receiving subcutaneous glatiramer acetate staying in the same health state, which means that the QALYs gained due to treatment were gained only due to delayed disease progression.

2.3.1 EQ-5D

The EQ-5D is a generic preference-based measure. It measures HR-QOL by asking patients how they feel today and scores health on five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The five dimensions have three answer categories, so the instrument can distinguish between 243 health states.

Fig. 1 Markov model structure. EDSS Expanded Disability Status Scale



Utility values are calculated by applying country-specific algorithms ('tariff') to the five-dimensional health profile. The tariff for this algorithm was developed by deriving utility values for 17 (Dutch tariff) and 42 (UK tariff) out of 243 health states in a valuation study with 298 (Dutch tariff) and 2997 (UK tariff) respondents from the general population using the time trade-off (TTO) method. The health state values of the states not included in the TTO study were estimated using regression techniques [13, 18]. The current study uses values from the three-level version of EQ-5D as the new five-level version was not included in the Multiple Sclerosis Risk Sharing Scheme Monitoring Study from which data was included. The range of the utility values of EQ-5D is 1 to -0.59 for the UK tariff [18] and 1 to -0.39 for the Dutch tariff [13].

2.3.2 29-Item Multiple Sclerosis Impact Scale (MSIS-29)

The MSIS-29 measures the physical and psychological impact of MS using a self-reported questionnaire consisting of 29 items [19]. The instrument is currently in its second version, which consists of the same 29 items as version 1 but has four answer categories rather than five, following extensive psychometric testing of the first version using item response theory [20]. Version 2 of the MSIS-29 has removed answer category 4 'quite a bit' from version 1. Both versions of MSIS-29 consist of two subscales, a physical impact scale (items 1–20) and a psychological impact scale (items 21–29).

For this study, utility algorithms based on both version 1 and version 2 are used, but only data for version 1 are available. Therefore, MSIS-29 version 1 was transformed to version 2 by merging the answer categories 'quite a bit' and 'moderately'. For version 1 to resemble version 2, the 'quite a bit' answers could be merged with either of the two adjacent levels 3 or 5. A Rasch analysis of the instrument by its developers showed that for both the physical impact scale and the psychological impact scale, it was hardest for respondents to distinguish between categories 2, 3 and 4 [20]. This suggests that it is more sensible to merge answer category 4 ('quite a bit') with answer category 3 ('moderately') than with category 5 ('extremely') when applying the algorithm for the MSIS-29 based on version 2 of the instrument.

The MSIS-29 has been converted to an instrument that can be used to calculate QALYs through mapping and direct valuation with TTO, as described below.

2.3.2.1 MSIS-29 Mapped on EQ-5D A simple search strategy was deployed in EMBASE and PubMed to identify mapping algorithms from MSIS-29 to EQ-5D using the search terms ['eq 5d' OR eq5d AND mapping AND 'msis 29' OR msis29] for EMBASE and [(eq-5D or eq5d) and

mapping and (msis-29 or msis29)] for PubMed. The EMBASE search identified 14 publications, of which two were mapping studies. The PubMed search identified two publications, which were the same as those identified in EMBASE: Versteegh et al. [10] and Hawton et al. [12]. The mapping by Versteegh et al. [10] included mapping to Dutch and UK EQ-5D utilities. The mapping by Hawton et al. [12] mapped to UK EQ-5D utilities. Hence, three published mapping algorithms were applied to estimate UK and Dutch EQ-5D utilities.

The mapping by Versteegh et al. [10] used data from the UK risk-sharing monitoring scheme to estimate both Dutch and UK EQ-5D utilities from MSIS-29 version 1. The ordinary least squares (OLS) algorithm was developed using a randomly drawn subsample of the UK risk-sharing monitoring scheme ($n = 661$) and was validated on two subsamples of the same dataset ($n = 339$ and $n = 295$). The final algorithms used nine (Dutch) and ten (UK) items of the MSIS-29 and a constant term to predict EQ-5D utilities, and did so without significant differences between observed and estimated values; however, subgroup analysis showed noticeable over-prediction of the utility values of patients in poor health.

The mapping by Hawton et al. [12] used data from the UK South West Impact of Multiple Sclerosis (SWIMS) Project, which contains version 2 of the MSIS-29. The estimation sample ($n = 672$) was used to test a range of mapping strategies. Predictive performance was subsequently tested on follow-up data across ten timepoints (2461 responses). The best-performing OLS model (model F in the original publication) includes eight MSIS-29 items and age and sex. To apply the Hawton et al. [12] mapping in this study, the MSIS-29 version 1 scores were re-scored to match the 4-point scale of version 2, by merging the categories 'quite a bit' and 'moderately' as described in Sect. 2.3.2.

2.3.2.2 Condition-Specific Multiple Sclerosis Impact Scale Preference-Based Measure (MSIS-PBM) (Dutch) The MSIS-PBM was derived from version 1 of the MSIS-29. A full description of the development of the measure can be found elsewhere [8]. Prior to valuation, the instrument was rescaled to a 4-point scale following Rasch analysis. Based on results of the Rasch analysis, psychometric properties and consultation of an expert neurologist, eight items of the MSIS-29 were used to describe 100 health states that were subsequently valued with the TTO method by a Dutch general population sample ($n = 402$). The values for health states not included in the TTO study were estimated using OLS regressions. The included items related to problems with balance, being clumsy, limitations in social and leisure activities at home, difficulties in using hands in everyday tasks, having to cut down on work or other daily

activities, feeling mentally fatigued, feeling irritable impatient or short tempered, and problems concentrating. The instrument has utility values ranging from 1 to 0.42.

2.3.2.3 Condition-Specific Eight-Dimensional Multiple Sclerosis Impact Scale (MSIS-8D) (UK) The MSIS-8D [6] was also derived from the MSIS-29, but from version 2. The MSIS-8D uses eight items of the MSIS-29 version 2 that take into account physical limitations, social and leisure time limitations, being stuck at home, having to cut down on work or daily activities, feeling mentally fatigued, feeling irritable, having problems concentrating and feeling depressed. Compared to MSIS-PBM it does not include problems with balance, being clumsy and difficulties in using hands. The MSIS-8D was converted to a preference-based instrument through a TTO study with 1702 respondents from the UK general population who valued 169 health states. The resulting instrument has utility values ranging from 0.882 to 0.08. The answer categories of the MSIS-29 version 1, which is included in the risk-sharing database, were re-scaled from 12,345 to 12,334 to match the design of the MSIS-8D, which is based on MSIS-29 version 2, similar to the approach of the Hawton et al. [12] mapping algorithm.

2.3.3 Comparison of MSIS-29 Items Included in the Utility Algorithms

Table 1 compares which items of the MSIS-29 (versions 1 and 2) have been included in the algorithms used to calculate utility values. It shows that all algorithms have included items from both the physical and the psychological subscale of the MSIS-29, but that none of the instruments are exactly similar despite similar development methods for both the condition-specific instruments and the mappings. This could be due to the use of different datasets and different versions of MSIS-29, expert advice or choices of the researchers during the development process.

2.4 Analyses

Equivalence of the three sources of utility values incorporated in the model were compared using paired t tests. The results of the Markov model are reported in terms of 2005 costs, QALYs and ICERs. The model is run using the literature-based utility values of the original Markov model in order to enable comparison of results of the replicated and original model. The effect of applying EQ-5D, mapped EQ5D and MSIS-PBM and MSIS8D utilities is then presented, for which the deterministic results are supplemented with a probabilistic sensitivity analysis.

2.4.1 Probabilistic Sensitivity Analysis

Probabilistic sensitivity analyses were not part of the original published model but were added for this study. Mean ICERS were calculated following probabilistic sensitivity analysis with 5000 Monte-Carlo simulations. Distributions were assigned to the costs of symptom management and to the utilities. Costs were varied using a gamma distribution and an assumed SE of 5 % of the mean cost value. Utilities were varied using a beta distribution and observed SEs in the UK risk-sharing scheme monitoring study.

2.4.2 Scenario Analyses

Three scenario analyses were run. The first scenario analysis tested the hypothesis that the findings of this study do not hold if the HR-QOL in the worst health state is as poor as identified by EQ-5D. If this hypothesis holds, it shows that the tendency of both regression-based mapping algorithms and condition-specific instruments to overestimate the HR-QOL of patients in poor health is a key driver of inflated ICERs. The hypothesis was tested by setting the utility value of health state EDSS4 equal to that of EQ-5D. The second scenario tested whether alternative recoding of the MSIS-29 version 1 to the MSIS-29 version 2, required for the application of the MSIS-8D tariff and the Hawton et al. [12] mapping algorithm, altered findings. In one alternative, the ‘quite a bit’ category was merged with the ‘extremely’ category, rather than merging the ‘quite a bit’ category with the ‘moderate’ category which would have been the most logical merge following Rasch analysis of the MSIS-29. As a last scenario analysis, an alternative mapping model by Hawton et al. [12] from the same publication was applied (mapping model B).

3 Results

The EQ-5D-, mapped EQ-5D- and condition-specific (MSIS-PBM and MSIS-8D)-based utility values per EDSS health state as derived from the risk-sharing monitoring scheme are presented in Table 2, alongside the utility values applied in the original Markov model. In the paired t test comparing Dutch mapped and observed EQ-5D values, values differed significantly for EDSS3 ($t = 2.1$, $p = 0.03$) and EDSS4 ($t = 2.7$, $p = 0.02$). Observed Dutch EQ-5D and MSIS-PBM values differed significantly for all health states ($t > 3.6$, $p < 0.05$) except for EDSS1.

Looking at the UK values, EQ-5D health state values estimated with the Versteegh et al. [10] mapping algorithm did not differ significantly from observed EQ-5D health states, except for EDSS4 ($t = 2.1$, $p = 0.05$). For the

Table 1 29-Item Multiple Sclerosis Impact Scale items included in utility algorithms

Item #	MSIS-29 items	MSIS-8D [6]	MSIS-PBM [8]	Versteegh et al. mapping [10]	Versteegh et al. mapping [10]	Hawton et al. mapping [12]
Country		UK	Dutch	Dutch	UK	UK
Version of MSIS-29		Version 2	Version 1	Version 1	Version 1	Version 2
<i>Physical impact scale</i>						
1	Do physically demanding tasks	×				
2	Grip things tightly					×
3	Carry things			×	×	
4	Problems with your balance		×			×
5	Difficulties moving about indoors			×	×	
6	Being clumsy		×	×	×	
7	Stiffness					×
8	Heavy arms and/or legs					
9	Tremor of your arms or legs					×
10	Spasms in your limbs			×	×	×
11	Your body not doing what you want it to do					
12	Having to depend on others to do things for you					
13	Limitations in social and leisure activities at home	×	×			
14	Being stuck at home more than you would like	×				
15	Difficulties using your hands in everyday tasks		×	×	×	
16	Having to cut down time spent on work/daily activities	×	×		×	
17	Problems using transport					×
18	Taking longer to do things				×	
19	Difficulty doing things spontaneously					
20	Needing to go to the toilet urgently					
<i>Psychological impact scale</i>						
21	Feeling unwell			×	×	×
22	Problems sleeping			×		
23	Feeling mentally fatigued	×	×			
24	Worries about your MS					×
25	Feeling anxious or tense					
26	Feeling irritable, impatient or short-tempered	×	×		×	
27	Problems concentrating	×	×			
28	Lack of confidence			×		
29	Feeling depressed	×		×	×	
Total number of MSIS-29 items included		8	8	9	10	8

MSIS-8D eight-dimensional Multiple Sclerosis Impact Scale, MSIS-29 29-item Multiple Sclerosis Impact Scale, MSIS-PBM Multiple Sclerosis Impact Scale Preference-Based Measure

Hawton et al. [12] model F, only EDSS4 was not significantly different ($t = 0.6$, $p = 0.55$). The condition-specific MSIS-8D values differed significantly for UK EQ-5D values ($t > 3.2$, $p < 0.01$) for all health states except EDSS2 ($t = 0.27$, $p = 0.78$).

Compared with symptom management, the incremental costs of subcutaneous glatiramer acetate were US\$25,277. When populated with the utility values in the original study, subcutaneous glatiramer acetate gained 0.13 QALYs and had an associated ICER of US\$195,065 compared with

Table 2 Utilities

	EDDS1	EDDS2	EDDS3	EDDS4	Relapse 1	Relapse 2
Bell et al. [16] values	0.82	0.68	0.53	0.49	0.73	0.59
Dutch utilities						
<i>N</i>	369	554	278	22		
EQ-5D-3L utilities	0.74	0.63	0.52	0.33	0.65	0.55
SE	0.01	0.01	0.02	0.06		
Mapped EQ-5D Versteegh et al. [10]	0.74	0.63	0.55	0.47	0.65	0.54
SE	0.01	0.01	0.01	0.03		
MSIS-PBM utilities	0.75	0.66	0.59	0.55	0.65	0.57
SE	0.01	0.01	0.01	0.02		
UK utilities						
<i>N</i>	367	555	278	21		
EQ-5D-3L utilities	0.71	0.58	0.44	0.19	0.62	0.49
SE	0.01	0.01	0.02	0.09		
Mapped EQ-5D Versteegh et al. [10]	0.71	0.58	0.46	0.33	0.62	0.49
SE	0.01	0.01	0.01	0.05		
Mapped EQ-5D Hawton et al. [12]	0.68	0.55	0.41	0.23	0.57	0.46
SE	0.01	0.01	0.01	0.03		
MSIS-8D utilities	0.68	0.58	0.51	0.42	0.59	0.49
SE	0.01	0.01	0.01	0.04		

Bold indicates a significant difference from EQ-5D at $p < 0.05$

EDSS Expanded Disability Status Scale, MSIS-8D eight-dimensional Multiple Sclerosis Impact Scale, MSIS-PBM Multiple Sclerosis Impact Scale Preference-based Measure, SE standard error

incremental costs of US\$57,174, 0.22 QALYs gained and an ICER of US\$258,465 in the original publication, indicating differences in mortality that could not be replicated from the published data.

When QALYs were computed with data from EQ-5D, mapped EQ-5D and the condition-specific instrument, the following results were found: 0.14 QALYs were gained when computed with Dutch EQ-5D utility data (0.16 with UK utilities), 0.10 with mapped Dutch EQ-5D (0.13 with UK Versteegh et al. [10] mapping and 0.14 with Hawton et al. [12] mapping) and 0.09 QALYs were gained when computed with utility data from the MSIS-PBM (0.095 with UK MSIS-8D). Deterministic and probabilistic ICERs for all instruments are reported in Table 3. Excluding productivity costs in the UK models did not alter the outcome.

The mean ICER after probabilistic sensitivity analysis using Dutch values was US\$180,783 (SE = 21,376) with EQ-5D-based values (US\$147,306 [SE = 17,597] with UK values), US\$246,071 (SE = 22,636) with mapped EQ-5D values (US\$193,327 [SE = 17,401] with mapped Versteegh et al. [10] UK values and US\$169,383 [SE = 14,412] with Hawton et al. [12] mapping) and US\$289,814 (SE = 20,115) with MSIS-PBM-based values (US\$265,372 [SE = 27,591] with MSIS-8D). The 5000 ICERS of the Monte-Carlo simulation for both the UK and the Dutch results are plotted in Fig. 2a, b and the associated

resulting cost-effectiveness acceptability curves for both the UK and The Netherlands are presented in Fig. 3a, b.

The hypothesis of the first scenario analysis was confirmed in the Dutch model: by setting the value of the worst health state of the mapped utilities and the condition-specific utilities equal to that of EQ-5D, the difference between the ICERs was reduced to US\$7000. In the UK model, the MSIS-8D-based ICER was still approximately US\$20,000 higher than the mapped and EQ-5D-based ICERs. The second scenario analysis, with the alternative coding of MSIS-29, increased the ICER, suggesting that the alternative coding did not alter the conclusion of this study. Applying mapping model B rather than model F increased the ICER and hence did not alter the conclusion of this study.

4 Discussion

Mapped and condition-specific utility values are considered to be a valid alternative to EQ-5D values in technology appraisals in the UK and The Netherlands. While authorities such as NICE request that they be informed about the effect of using these alternatives on the ICER, such a comparison is often not possible as the alternative methods can only be applied when EQ-5D is either absent or not appropriate. This study used EQ-5D (Dutch and UK

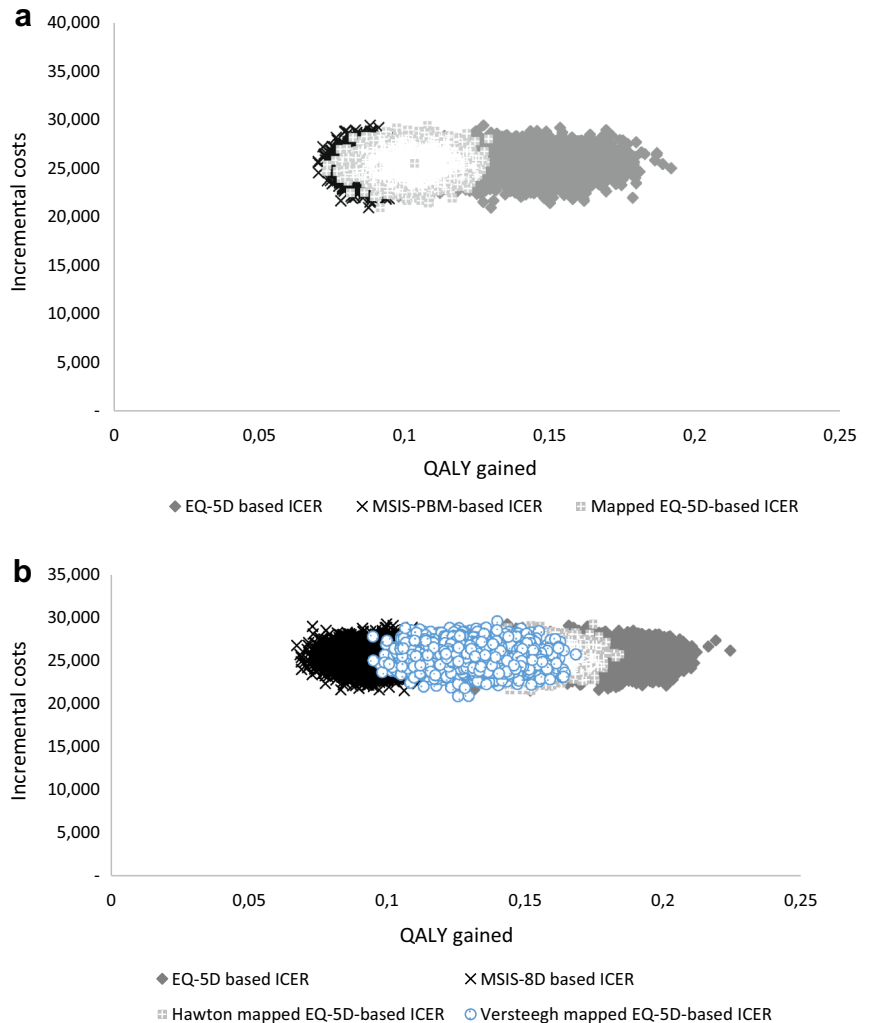
Table 3 Deterministic and probabilistic incremental cost-effectiveness ratios for symptom management versus glatiramer acetate

	Incremental costs (\$US)	Incremental QALYs ^a	Deterministic ICER (\$US)	Probabilistic ICER (\$US)	% difference vs. EQ-5D
Dutch utilities					
EQ-5D-3L utilities	25,277	0.14	182,291	180,783	
Mapped EQ-5D Versteegh et al. [10]	25,277	0.1	245,363	246,071	36
MSIS-PBM utilities	25,277	0.09	289,514	289,814	60
UK utilities					
EQ-5D-3L utilities	25,277	0.16	153,476	140,736	
Mapped EQ-5D Hawton et al. [12]	25,277	0.14	170,771	169,383	20
Mapped EQ-5D Versteegh et al. [10]	25,277	0.13	194,445	193,327	37
MSIS-8D utilities	25,277	0.1	265,342	265,372	89

ICER incremental cost-effectiveness ratio, MSIS-8D eight-dimensional Multiple Sclerosis Impact Scale, MSIS-PBM Multiple Sclerosis Impact Scale Preference-based Measure, QALYs quality-adjusted life-years

^a Rounded values

Fig. 2 Cost-effectiveness planes: **a** Dutch and **b** UK. ICER incremental cost-effectiveness analysis, MSIS-8D eight-dimensional Multiple Sclerosis Impact Scale, MSIS-PBM Multiple Sclerosis Impact Scale Preference-Based Measure, QALY quality-adjusted life-year



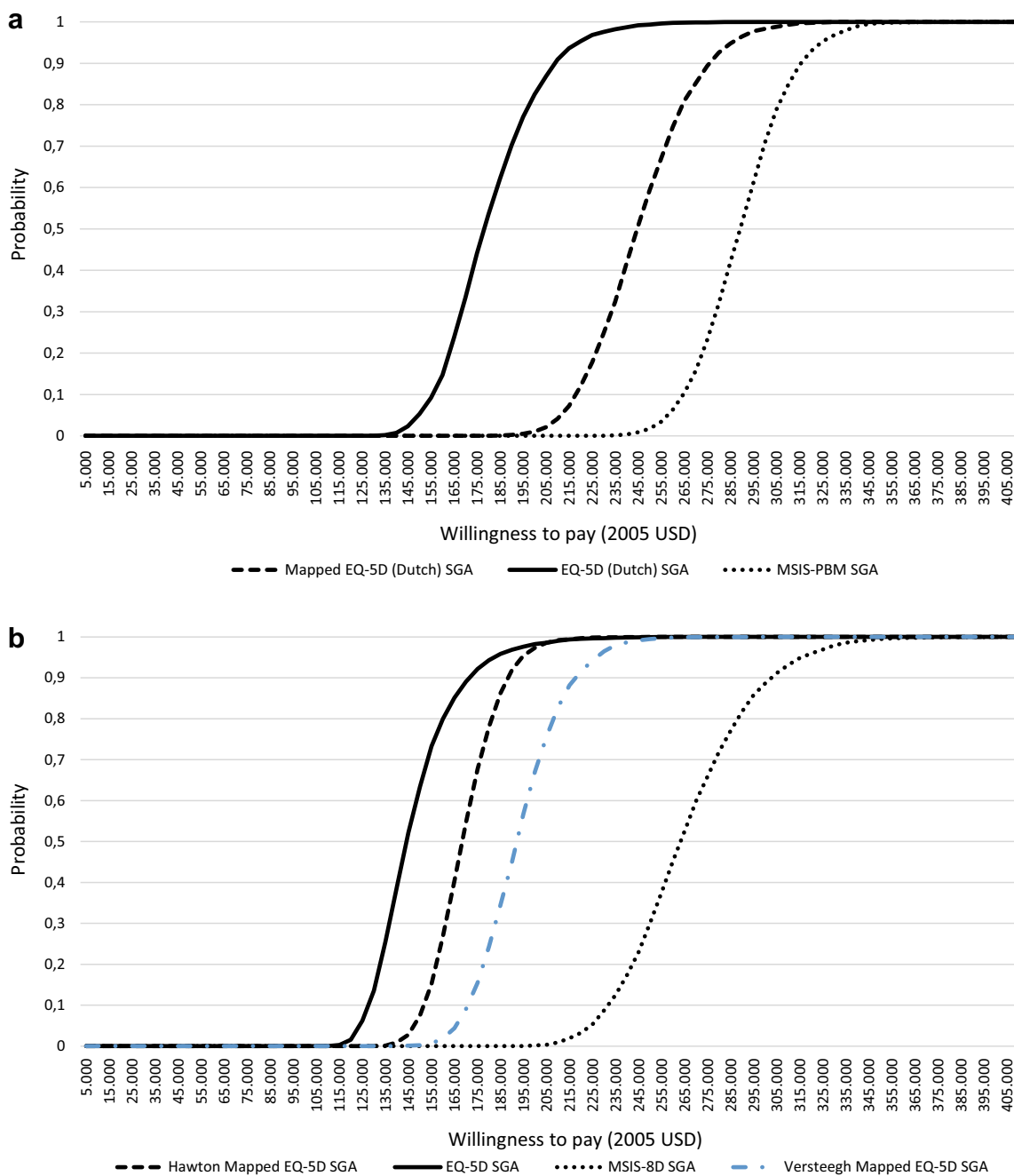


Fig. 3 Cost-effectiveness acceptability curves: **a** Dutch and **b** UK. *MSIS-8D* eight-dimensional Multiple Sclerosis Impact Scale, *MSIS-PBM* Multiple Sclerosis Impact Scale Preference-Based Measure, *SGA* subcutaneous glatiramer acetate

tariff), mapped EQ-5D (Dutch and UK mapping models) and condition-specific utility values (Dutch and UK instruments) taken from the same patient sample in a hypothetical Markov model for relapsing-remitting MS patients and showed that the ICERs differed substantially when applying alternative sources of utilities.

This study was conducted in the context of the UK and Dutch guidelines for economic evaluations, which both prefer a uniform assessment of HR-QOL with EQ-5D

[2, 4]. When EQ-5D is demonstrably inappropriate, condition-specific-utility instruments are accepted by the UK and Dutch guidelines. The condition-specific instruments applied here include dimensions not included in EQ-5D (fatigue and cognition) and have advantages (increased discriminatory properties) and disadvantages (floor effect for MSIS-PBM and floor and ceiling effect for MSIS-8D compared to EQ-5D). The consequence of these disadvantages is demonstrated in this study through a

remarkably higher ICER of the condition-specific instruments. This finding highlights the importance of having strict criteria for determining the ‘inappropriateness’ of EQ-5D in the context of health economic modelling that places the most emphasis on the absolute difference in utility between health states. The EQ-5D instrument may seem inappropriate and lack face validity in some conditions, or even have unfavourable psychometric properties, but condition-specific instruments may underestimate the total QALY gain due to a reduced scope of severity. The results of this study also confirm the specification of the guidelines that mapping to EQ-5D is preferred over condition-specific alternatives in the absence of EQ-5D, as the ICERs based on mapped EQ-5D better compare with ICERs based on EQ-5D and, hence, contribute to a uniform assessment of benefit.

The foremost explanation for the increase of the ICER when using mapped and condition-specific utility values is the smaller absolute difference between utility values of the different health states in the Markov model, particularly for the two poorer health states EDSS3 and EDSS4, for which mapped and condition-specific utility values were higher than EQ-5D. The scenario analysis showed that the higher utility values in the EDSS4 state are a main driver of the difference between ICERs. This shows that the over-prediction of patients in poor health by mapping models, and the generally higher values of condition-specific utility instruments, have a large impact on technology appraisals in the UK and The Netherlands.

Over-prediction of low utility values in OLS mapping models is frequently observed [21, 22] and its impact on the observed QALY gain has also been described [23]. Here it is shown once more that the inability of mappings based on OLS regression algorithms to adequately capture the utility of patients in poor health is problematic and can even cause incorrect decisions in health technology appraisals: mapped utilities may suggest that health would be displaced with a positive reimbursement decision, while this is merely an artefact of an unsuccessful estimation method.

It was expected that the condition-specific MSIS-PBM and MSIS-8D would be better at identifying improvements in quality of life than EQ-5D. MSIS-PBM has been demonstrated previously to be more sensitive to change than EQ-5D in EDSS categories 3–5 [8] and did not demonstrate a ceiling effect compared to EQ-5D. However, as mentioned in Sects. 1 and 2, this improved sensitivity is accompanied by a floor effect of both measures compared to EQ-5D. For example, while the lowest utility value of MSIS-8D is lower than that of MSIS-PBM (0.08 vs. 0.42), not many patients actually reach such a state, explaining the high mean value of patients in poor health. For MSIS-PBM, the largest total utility decrement for the very ill is

0.58 (and, hence, a theoretical minimum utility value of 0.42) compared with 1.329 for the Dutch tariff of EQ-5D [13]. As a consequence, the value for the poorest EDSS health states is rather high. This high floor value is also observed in the SF-6D and in several other condition-specific preference-based measures [24–26] and might be particularly influential in lifetime models where patients deteriorate over time and hence reach a health state associated with the floor value. While the MSIS-PBM and the MSIS-8D might have smaller variance than EQ-5D, demonstrated in the size of the ICER cloud after probabilistic sensitivity analysis in Fig. 2a, b, it is the absolute differences that have the largest impact on the ICER.

The observation that mapped and condition-specific utility values inflate the ICER is not generalisable to technology appraisals of life-saving/-extending interventions or to Markov models with very short time horizons. If an intervention would prolong life, and the mapped and condition-specific utility values are higher on average than EQ-5D utility values for patients in poor health, the additional life-years are assigned a larger QALY gain. The long-term nature of MS, combined with no effect of medication on survival, increase the difference between the seven ICERs calculated for this study. If a model has a very short time horizon, such as in cost-effectiveness analyses performed alongside clinical trials, not all patients reach the very poor EDSS states, decreasing the difference between EQ-5D, mapped EQ-5D and condition-specific utilities. As such, it could be that the effects here would not be replicated in economic evaluations alongside a clinical trial. It is even conceivable that a trial-based economic evaluation with a short time horizon would produce results opposite to those observed here as such an evaluation does not require the grouping of patients in Markov health states. However, it should be noted that guidelines call for economic evaluations that are able to capture all relevant costs and effects of a treatment to a patient, which generally requires time horizons that are much longer than the trial period, especially for long-term conditions such as MS. Therefore, the effects observed here are likely to be observed in economic evaluations of MS that are acceptable to reimbursement authorities.

When compared with the utility estimates applied in this study, the utility estimates of the original study are high and resemble MSIS-PBM utilities, as shown in Table 2. The utilities in the original Bell et al. [16] study were taken from the literature by the authors of the study. The first cited source in that study is Prosser et al. [27], which applied the standard gamble method in 67 members of the general public of the USA to value six MS health states with ten attributes, of which the worst health state was related to EDSS score 8. The health state that represented EDSS score 8 had a mean utility score of 0.491. The second

cited study was a study report by Kobelt et al. [28], who used the three-level version of EQ-5D in combination with a non-final version of the US tariff as well as the UK tariff in 1909 MS patients. Utilities were reported in three categories: mild, moderate and severe, of which the severe category consists of patients with an EDSS score >6.5 . This category had a mean utility of 0.698 with the US tariff and 0.368 with the UK tariff. The lowest possible value of the US tariff is -0.109 [29]. The study noted an under-representation of EDSS category 8 and above. The third cited study did not contain utility values [30]. Standard gamble utilities are generally higher than TTO-based utilities due to probability weighting, loss aversion and scale compatibility [31]. Given these sources of the utility values for the literature-based model, the difference from the utility values used here based on the TTO-valued EQ-5D and MSIS-PBM is understandable. The ICER using the utility values from the literature was in between the EQ-5D-based ICER and the mapped and MSIS-PBM-based ICER, mostly because of the larger range of the utility values, i.e. the difference between the best health state and the worst health state in the model, which allows a larger effect.

A limitation of this study is that the Markov model from the literature was not fully replicated as not all information could be reproduced from the published article. Contact with the authors from the original paper was sought on separate occasions but not established. While the ICER is rather similar to that of the replicated study, the difference between the total QALYs gained and total incremental costs indicate a systematic difference between the original model and the one replicated for this study. A crucial variable seems to be total mortality, which for both studies was taken from 2005 US life tables and a 0.000952 monthly probability of progressing from EDSS4 to death. The key difference between this study and the published study was a fixed effectiveness rate rather than long-term prediction models. In the original study, disease progression increased over time, causing patients to progress faster to the third and fourth EDSS health state in which costs and the probability of death are larger. Also, the transition probabilities of relapse rates and drug use in relapse states could not be fully replicated from the original publication and if the interpretation was not correct, this will also impact the ICER. Hence, it is of importance to stress the hypothetical nature of the absolute value of the ICERs produced in this study. However, the relative rather than absolute difference between the MSIS-PBM-based ICER and the EQ-5D-based ICER is unlikely to be affected by these parameters.

Another limitation of this study is that the Markov model was based on EDSS health states, and only on four merged categories, rather than, for example, a separate health state for all EDSS scores. First, a Markov model based on EDSS health states is inherently limited to the discriminatory

properties of EDSS, which are arguably limited as it is a mainly mobility-based assessment of quality of life. Second, using only a limited number of EDSS states (effectively grouping all EDSS categories into four states) reduces the potential for the condition-specific instruments to show increased sensitivity. For example, as measured with MSIS-PBM, the difference in average utility values between EDSS categories 3 and 5 is 0.05 with MSIS-PBM, while EQ-5D is not able to pick up that improvement. This means that, had a Markov model been used with all EDSS states, the MSIS-PBM would be able to capture the benefit of a drug that delays patients' transitions between EDSS categories 3, 4 and 5. It is, however, unlikely that such a design would alter the conclusions of this study. The additional benefit that could be captured between EDSS categories 3 and 5 is a utility of 0.05, which is much smaller than the difference between the utility values of the fourth and worst EDSS Markov health state (which is 0.14). This means that the improved sensitivity of the instrument has a smaller impact on the outcome of economic evaluations than its reduced scope of severity. Hence, it is unlikely to offset the difference in the ICERs between EQ-5D and MSIS-PBM, even when the model had had a separate state for all EDSS categories to best capture the increased sensitivity of the condition-specific instruments. In general, it would be important to see if the results demonstrated here are replicated in other health economic models for MS.

The two condition-specific instruments in MS applied here have disadvantages that seem to outweigh their increased sensitivity, at least in the context of economic evaluations in MS with a lifetime time horizon. Regarding the future of HR-QOL assessment in MS, a future avenue of research that may be fruitful is adding relevant MS dimensions (i.e. fatigue) to the descriptive system of EQ-5D and assigning a specific utility decrement to those dimensions.

5 Conclusion

The overestimation of quality of life of patients with MS by mapped EQ-5D or condition-specific utility values, relative to observed EQ-5D, increases the ICER substantially in a Markov model with a lifetime time horizon for MS.

Acknowledgments The author would like to thank Saskia de Groot, Laura Burgers and Pieter van Baal of the institute for Medical Technology Assessment (iMTA) of Erasmus University of Rotterdam for sharing their suggestions for this study.

Compliance with Ethical Standards

The research was not externally funded.

Disclosure Matthijs Versteegh is a member of the EuroQoL Research Foundation.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Mackenzie IS, Morant SV, Bloomfield GA, MacDonald TM, O'Riordan J. Incidence and prevalence of multiple sclerosis in the UK 1990–2010: a descriptive study in the General Practice Research Database. *J Neurol Neurosurg Psychiatry*. 2014;85(1):76–84.
- National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. London: NICE; 2013. <http://publications.nice.org.uk/pmg9>. Accessed 7 Jan 2016.
- Brazier J, Rowen D. NICE DSU technical support document 11: alternatives to EQ-5D for generating health state utility values. Report by the Decision Support Unit. Sheffield: Decision Support Unit; 2011. http://www.nicedsu.org.uk/TSD11%20Alternatives%20to%20EQ-5D_final.pdf. Accessed 1 June 2016.
- Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Diemen: Zorginstituut Nederland; 2015.
- Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? *Health Econ*. 2010;19(2):125–9.
- Goodwin E, Green C, Spencer A. Estimating a preference-based index for an eight-dimensional health state classification system for multiple sclerosis. *Value Health*. 2015;18(8):1025–36.
- Kuspinar A, Mayo NE. Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? *Health Qual Life Outcomes*. 2013;11:71.
- Versteegh M, Leunis A, Uyl-de Groot C, Stolk E. Condition-specific preference-based measures: benefit or burden? *Value Health*. 2012;15:504–13.
- Orme M, Kerrigan J, Tyas D, Russell N, Nixon R. The effect of disease, functional status, and relapses on the utility of people with multiple sclerosis in the UK. *Value Health*. 2007;10(1):54–60.
- Versteegh MM, Leunis A, Luime JJ, Boggild M, Uyl-de Groot CA, Stolk EA. Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. *Med Decis Mak*. 2012;32(4):554–68.
- Kobelt G, Berg J, Lindgren P, Fredrikson S, Jonsson B. Costs and quality of life of patients with multiple sclerosis in Europe. *J Neurol Neurosurg Psychiatry*. 2006;77(8):918–26.
- 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 Health*. 2012;15(8):1084–91.
- Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJV. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ*. 2006;15(10):1121–32.
- Kobelt G, Jönsson L, Henriksson F, Fredrikson S, Jönsson B. Cost-utility analysis of interferon beta-1b in secondary progressive multiple sclerosis. *Int J Technol Assess Health Care*. 2000;16(3):768–80.
- Nuijten MJ, Hutton J. Cost-Effectiveness Analysis of interferon beta in multiple sclerosis: a Markov process analysis. *Value Health*. 2002;5(1):44–54.
- Bell C, Graham J, Earnshaw S, Oleen-Burkey M, Castelli-Haley J, Johnson K. Cost-effectiveness of four immunomodulatory therapies for relapsing-remitting multiple sclerosis: a Markov model based on long-term clinical data. *J Manag Care Pharm*. 2007;13(3):245–61.
- Boggild M, Palace J, Barton P, Ben-Shlomo Y, Bregenzer T, Dobson C, et al. Multiple sclerosis risk sharing scheme: two year results of clinical cohort study with historical comparator. *BMJ*. 2009;339:b4677.
- Dolan P. Modeling valuations for the EuroQol health states. *Med Care*. 1997;35:1095–108.
- Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. *Brain*. 2001;124(5):962–73.
- Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;13(12):iii, ix–x, 1–177.
- Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*. 2010;11(2):215–25.
- Versteegh MM, Rowen D, Brazier JE, Stolk EA. Mapping onto Eq-5 D for patients in poor health. *Health Qual Life Outcomes*. 2010;26(8):141.
- Barton GR, Sach TH, Jenkinson C, Avery AJ, Doherty M, Muir KR. Do estimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? *Health Qual Life Outcomes*. 2008;14(6):51.
- Stolk EA, Busschbach JJV. Validity and feasibility of the use of condition-specific outcome measures in economic evaluation. *Quality Life Res*. 2003;12(4):363–71.
- Yang Y, Brazier JE, Tsuchiya A, Young TA. Estimating a preference-based index for a 5-dimensional health state classification for asthma derived from the asthma quality of life questionnaire. *Med Decis Mak*. 2011;31(2):281–91.
- Brazier JE, Roberts J, Platts M, Zoellner YF. Estimating a preference-based index for a menopause specific health quality of life questionnaire. *Health Qual Life Outcomes*. 2005;15(3):13.
- Prosser LA, Kuntz KM, Bar-Or A, Weinstein MC. Patient and community preferences for treatments and health states in multiple sclerosis. *Mult Scler*. 2003;9(3):311–9.
- Kobelt G, Berg J, Atherley D, Hadjimichael O. Costs and quality of life in multiple sclerosis: a cross-sectional study in the United States. *Neurology*. 2006;66(11):1696–702.
- Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43(3):203–20.
- Kobelt G. Economic evidence in multiple sclerosis: a review. *Eur J Health Econ*. 2004;5:s54–62.
- Bleichrodt H. A new explanation for the difference between time tradeoff utilities and standard gamble utilities. *Health Econ*. 2002;11:447–56.