Article:

Rowen, D. orcid.org/0000-0003-3018-5109, Powell, P., Mukuria, C. et al. (3 more authors) (2021) Deriving a preference-based measure for people with Duchenne Muscular Dystrophy from the DMD-QoL. Value in Health, 24 (10). pp. 1499-1510. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2021.03.007

ScienceDirect

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

Deriving a Preference-Based Measure for People With Duchenne Muscular Dystrophy From the DMD-QoL

Donna Rowen, PhD, Philip Powell, PhD, Clara Mukuria, PhD, Jill Carlton, PhD, Richard Norman, PhD, John Brazier, PhD

ABSTRACT

Objectives: This study generates a preference-based measure for capturing the quality of life of people with Duchenne muscular dystrophy (DMD) from a new measure of quality of life, DMD-QoL.

Methods: A health state classification system was derived from the DMD-QoL based on psychometric performance of items, factor analysis, and item response theory analysis. Preferences for health states described by the classification system were elicited using an online discrete choice experiment survey with life years as an additional attribute, from members of the UK general population (n = 1043). Discrete choice experiment data was modeled using a conditional fixed-effects logit model and utility estimates were directly anchored on the 1 to 0 full health-dead scale.

Results: The health state classification system has 8 dimensions: mobility, difficulty using hands, difficulty breathing, pain, tiredness, worry, participation, and feeling good about yourself. The standard model had mostly statistically significant coefficients and reflected the instrument's monotonic structure. However, 2 dimensions had inconsistent coefficients (where utility increased as health worsened) and a consistent model was estimated that merged adjacent inconsistent severity levels. The best state defined by the classification system has a value of 1 and the worst state has a value of −0.559.

Conclusion: The modeled results enable DMD-QoL-8D utility values to be generated using DMD-QoL or DMD-QoL-8D data to generate QALYs for people with DMD. QALYs can then be used to inform economic models of the cost-effectiveness of interventions in DMD. Future research comparing the psychometric performance of DMD-QoL-8D to existing generic preference-based measures, including EQ-5D-5L, is recommended.

Keywords: condition-specific preference based measure, DMD-QoL, Duchenne muscular dystrophy, utilities, QALYs.

VALUE HEALTH. 2021; ■(■):■–■

Introduction

Cost-effectiveness analysis is an important tool internationally for assessing whether new or existing healthcare interventions are cost-effective, by assessing incremental costs and benefits. Benefits of treatments can be captured using quality adjusted life years (QALYs), which capture both quality and quantity of life. Quality of life (QoL) is measured by assigning a health state utility value on a 1 to 0 full health-dead scale, and this is multiplied with duration in the health state to generate the QALY. Health state utility values are typically measured using generic preference-based measures (PBM), following recommendations by international agencies on how benefits should be measured for assessing new interventions.[1,2] However, generic PBM are not always available, appropriate, or sufficient, and there is a role for other ways of measuring utilities that may be more accurate for a particular condition or population. In these circumstances, condition-specific PBM are one option for generating health state utility values, which are typically developed from existing validated patient reported outcome measures (PROM). Condition-specific PBM are often patient focused and more relevant for patients than generic PBM and better able to focus on condition-specific and symptomatic aspects that are relevant for a particular condition or population. However, the use of condition-specific PBM may impact on the comparability and generalizability of QALYs across different conditions.

Duchenne muscular dystrophy (DMD) is a rare inherited condition that predominantly affects boys. The disease has an estimated incidence of around 1:5000 live male births.[3] Physical and functional problems are manifest in people with DMD from as early as 2 years[4] due to an absence of the dystrophin protein, which causes muscle damage and progressive weakness. Gradual impairments in physical functioning occur over time, eventually resulting in cardiovascular and respiratory problems. The life expectancy of people with DMD is shortened as a result, at a median of 30 years,[5] yet advances in healthcare are progressively improving this estimate. There is a large impact on caregivers of people with DMD, including financial burden.[6] There is no known curative treatment for DMD, and clinicians aim to treat symptoms and slow progression and, in doing so, improve patients' QoL. PROM are typically used to assess QoL. Research using PROM to assess QoL in DMD has reported that it is lower than general population comparison groups,[7] including in the domains of physical and social functioning.[8,9] These findings notwithstanding, concerns have been raised about the capacity of available PROM and PBM to optimally measure QoL in DMD.[10,11]

For a PROM to be useful, it is essential that it has a high reliability and validity to measure the construct of interest (QoL) in the patient group of interest (DMD). A key element of the validity of a PROM is content validity, which includes comprehensiveness, or that all the key elements of QoL that are important to patients are assessed.[12] A recent review catalogued a number of relevant QoL themes identified in patients with DMD, such as participation and accessibility,[7] which may not be optimally assessed by available PROM and PBM. A complementary review evaluating the content of PROM used to assess QoL in DMD, using Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) methods,[12] concluded that no PROM had high quality evidence for its content validity in DMD.[13] Measuring QoL in DMD is further complicated by the typical use of different PROM across the life course, impeding comparisons across time, patients, and disease stage, which could be remedied with a common instrument.

The DMD-QoL was recently developed as a condition-specific PROM to address some of the problems identified with existing measures and provide a content valid measure of QoL in people with DMD from childhood to adulthood.[14] An aim in the development of the DMD-QoL was to produce a PROM that could be adapted into a PBM for use in cost utility analysis.[15] Alternative, available generic PBM, such as EQ-5D, were evaluated as having poor content validity for DMD,[13] although these have been previously used to generate utility estimates.[16] Previous QoL questionnaires designed for use specifically in people with muscular dystrophy have not been designed for use in economic analysis, and thus cannot be used in cost utility calculations directly.

This study generates a DMD-specific PBM to produce health state utility values for the DMD-QoL. A PBM consists of (1) a classification system used to describe the QoL of a respondent through assigning them to a health state, and (2) a scoring system used to generate utility values for all health states defined by the classification system on the 1 to 0 full health-dead scale used to generate QALYs.

The derivation of the DMD-specific PBM involved 3 stages. First, a classification system was derived from the DMD-QoL. Second, an online discrete choice experiment (DCE) was conducted to elicit preferences. Third, the DCE data was modeled to generate utility values for all health states defined by the classification system. The DMD-specific PBM enables utility values to be generated from existing DMD-QoL data. These utility values can then be used to estimate QALYs for use in cost-effectiveness analyses of interventions for DMD, for submission to agencies such as NICE (National Institute of Health and Care Excellence).[17]

Methods

DMD-QoL

The DMD-QoL is a new condition-specific PROM developed to measure QoL in people with DMD aged 7 and above.[14] The measure has 14 items covering 3 QoL dimensions (physical functioning, psychological impact, and social participation). Each item has 4 frequency levels that describe the severity of each attribute (ie, never, sometimes, a lot of the time, and all of the time).

The development of the DMD-QoL involved multiple stages that followed the process outlined in a published protocol[15] and included patient, caregiving, and clinician involvement throughout. The DMD-QoL was developed across multiple stages, involving qualitative interviews with people with DMD; face validation exercises with patients, caregivers, and clinicians; and analyses of psychometric data from an online patient survey. Patient and public involvement and engagement was embedded throughout development of the PROM.

Derivation of health state classification system for valuation

The first step in valuing the DMD-QoL was to establish a multidimensional health state classification system (HSC) that would be amendable to valuation using DCE. As valuation via DCE can be a cognitively demanding task, and because people can only typically process between 5 and 9 pieces of information,[18] it is common to limit the number of dimensions in an HSC. We chose to limit the HSC to 8 dimensions (plus duration), which is a number of dimensions that has been successfully used previously in the HSCs for other PBM (for example[19-21]).

Initial psychometric analyses of the DMD-QoL suggested a correlated 3-factor structure (physical functioning, psychological impact, and social participation). As the DMD-QoL has 14 items with a 3-factor structure, some prioritization across items was thus necessary to reduce the length of the questionnaire for valuation, while aiming to retain representation across the underlying 3 factors. Selection of DMD-QoL items for the HSC was determined by selecting ≥2 items for each underlying factor. Items were selected based on their perceived importance in the underlying qualitative and developmental work for the DMD-QoL[14] and performance of the items psychometrically and in item response theory (IRT) analyses (using a rating scale model), based on an approach[22] used to develop several PBM (for example[19,20,23]). Items were judged for selection based on available data using the following criteria: (1) show adequate item infit and outfit (based on mean squared residual values) of between 0.5 and 1.5[24]; (2) generate the largest spread (and thus discrimination) across each latent factor[22]); (3) maximize independence between items for use in valuation (for example see[25]), using the average intra-factor correlation (ie, average correlation with items within the same factor) based on a polychoric correlation matrix; (4) minimize floor and ceiling effects and missing data; and (5) minimize disordered thresholds and differential item functioning, if possible. Differential item functioning was assessed based on report method (self-report vs proxy report) and age (≥16 years or 7-15 years) at a 1% alpha level[26]). Finally, all items had to make sense for valuation, in combination with their proposed severity levels (eg, being embarrassed "all of the time" may be unrealistic).

Selection for the HSC relied on data produced from an online survey using a draft (27 item) version of the DMD-QoL, reported in detail elsewhere.[14] DMD-QoL data existed for a valid sample of 102 people with DMD (n = 37 self- or assisted self-reported and n = 66 proxy reported). All participants were male, and patients' ages ranged from 7 to 44 years, but were skewed toward younger ages (M = 15.77, SD = 7.87). A full description of the sample is available elsewhere.[14] All psychometric and IRT analyses were conducted in R x64 3.6.1,[27] utilising packages psych,[28] and lavaan[29] for the psychometrics and eRm[30] and lordif[26] for the IRT analyses.

Valuation technique

Discrete choice experiments are being increasingly used to generate health state utility values.[31,32] A DCE task for health state valuation typically involves a choice between 2 health profiles, where each profile is described using a selected level for each dimension and the duration of this health state followed by death. Participants are asked to select their preferred profile. The inclusion of duration as an attribute in the DCE alongside the QoL dimensions enables the modeled latent values from the DCE to be directly anchored onto the 1 to 0 full health-dead scale required for QALYs through the inclusion of an additional attribute, often called DCE_{TTO}.[33-35] This technique has been successfully used online in a number of valuation studies (for example[21,36-38]).

Selecting profiles for the DCE survey

The aim of the DCE was to generate utility values for the health state classification system using duration by providing participants with pairwise choices of the HSC (8 dimensions) and duration. The levels of the duration attribute were selected as 1, 4, 7, and 10 years in line with previous studies.[21,36-38] The 8 dimensions combined with duration would result in >260 000 health profiles (4^9 = 262 144), meaning all possible combinations of profiles could not be included in the DCE. A subset therefore needed to be selected to enable estimation of utility values for the HSC.

Although the DCE task is easy to understand, each pairwise choice would have 18 pieces of information (9 per health profile), which would increase the cognitive burden for participants making choices. We simplified the cognitive task by constraining the number of dimensions from the DMD-QoL that could vary in each pairwise choice to 4.[39,40] This means that for each pairwise choice, participants only saw differences in severity in 4 of the 8 out of the DMQ-QoL HSC and duration; each of the other 4 dimensions were the same.

Ngene 1.2.1[41] was used with a candidate set (n = 22 336) of random pairs that had the required constraints across 4 dimensions. The severity levels for each dimension were treated as categorical and duration was treated as continuous. The model included interaction terms between the levels for each dimension and duration to allow estimation of the associated disutility. To optimize and select the design, we maximize C-efficiency, which focuses on the ratio of coefficients, which will be used to estimate a value set. We used a willingness to pay approach derived using the modified Fedorov algorithm with small nonzero priors for the interaction terms to denote monotonically increasing severity in levels within each dimension.[41,42]

The number of choice tasks for DCEs with 2 alternatives needs to be at least as large as the number of parameters to be estimated.[43] A recent review suggested that the number of choice tasks in experiments ranged from 12 to 3160 but were typically around 150 to 200.[29] Therefore, we selected 180 choice sets to achieve both good coverage of the design space and a large enough number of observations per choice pair to allow estimation of probabilities for each.

Respondents were purely randomly allocated 9 of the 180 choice sets in the survey. To examine respondent engagement and understanding, all respondents also answered the same practice DCE task and the same dominance DCE task. Both the practice and dominance DCE tasks had one dominant health profile (ie, a health profile that had a better or same level across all attributes than the profile to which it was being compared). The 9 DCE choice sets that were randomly selected from the design and the dominance DCE task were randomly ordered for each respondent.

DCE survey

Respondents were recruited using an existing online panel from a market research agency, and quotas were set for age and sex to ensure a representative sample of the UK population in terms of age and sex according to the 2011 UK census. Members of the online panel received an invite to the survey in their agency portal and the survey remained open until all combined age and sex quotas for a sample size of 1000 were met. Survey respondents were thanked for their participation with a nominal number of vouchers that can be accumulated and exchanged for goods.

The survey began with an information sheet about the survey and informed consent was taken prior to starting the survey. The survey had 3 stages. First, respondents completed sociodemographic, health, and QoL questions. Respondents also completed the classification system to familiarize respondents with the descriptions of the health state profiles used in the survey. Participants were not informed that the classification system (or later health states) was for people with DMD, and DMD was not mentioned throughout the survey. Second, respondents completed one practice DCE question that explained the question and gave feedback about their choice, enabling respondents to amend their choice and complete the practice question again. Respondents then completed 10 DCE tasks, 9 were selected from the design and 1 was a dominance task (see Fig. 1 for an example of the DCE task). Third, respondents completed 2 questions about how difficult the DCE tasks were to understand and answer.

Yellow highlighting was used to indicate where the 2 profiles differed to make the task easier to understand for respondents, and this technique has been successfully used previously.[21,39] Prior to undertaking the main survey, the survey was soft launched with 100 participants and the data analyzed. Because no changes were made to the survey following the soft launch, this data is included in the final dataset. The research received ethical approval from the UK National Health Service (NHS; REC reference: 18/SW/0055).

Analysis

The sociodemographic and health characteristics of the sample were reported. The DCE_{TTO} data was analyzed using a widely-used model specification[33]:

$$(1)\mu_{ij} = \alpha_i + \beta_1 t_{ij} + \beta_2^{'} \mathbf{x}_{ij} t_{ij} + \varepsilon_{ij}$$

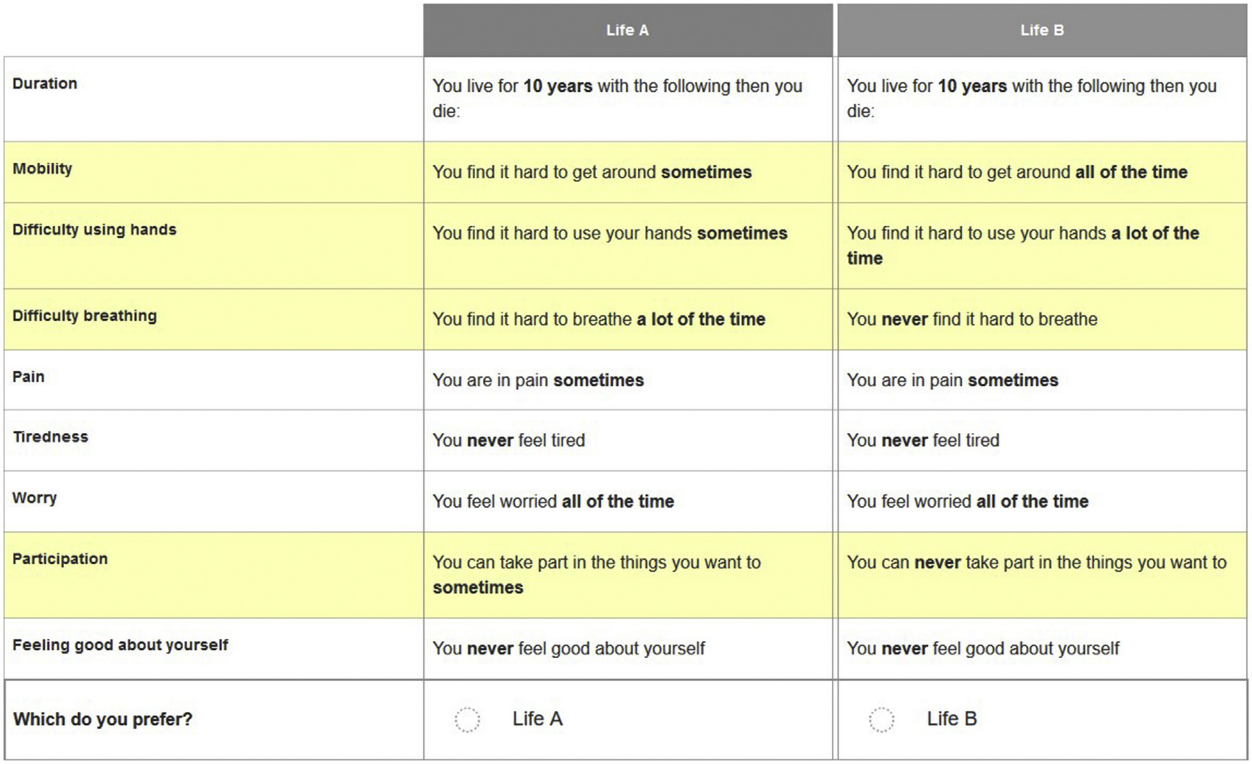
where μ_{ij} represents the utility of individual i for health profile j,$\alpha_i$is an individual specific constant term, ε_{ij} represents the error term, β_1 is the coefficient for duration in life years t and $\beta_2^{'}$ represents the coefficients on the 24 interaction terms of duration and severity levels of the dimensions comprised of levels 2, 3 and 4 (level 1 is the baseline). Using this specification, duration was modeled as a linear and continuous variable. This assumption was examined by modelling duration as a categorical variable and plotting the duration coefficients.[44]

Model estimation was undertaken in Stata version 15 using the conditional logit fixed effects model with cluster adjusted standard errors. Model performance was examined using the sign, significance and logical consistency of coefficients, log likelihood and pseudo R-squared.

The coefficients generated using this model specification are latent values that are not anchored onto the 1 to 0 full health-dead scale required to generate QALYs. The coefficients are anchored onto the 1 to 0 full health-dead scale using a widely used technique,[30] which uses the marginal rate of substitution, calculated by dividing the coefficient for each level γ of each attribute δ by the coefficient for duration, $\frac{\beta_{2\gamma\delta}}{\beta_1}$, and standard errors were calculated using the Delta method.

For the anchored results to be used to generate health state utility values, utility values must decrease or remain the same as health worsens. To ensure this a consistent model was estimated,

Figure 1. Survey screenshot of a discrete choice experiment task.



	Life A	Life B
Duration	You live for **10 years** with the following then you die:	You live for **10 years** with the following then you die:
Mobility	You find it hard to get around **sometimes**	You find it hard to get around **all of the time**
Difficulty using hands	You find it hard to use your hands **sometimes**	You find it hard to use your hands **a lot of the time**
Difficulty breathing	You find it hard to breathe **a lot of the time**	You **never** find it hard to breathe
Pain	You are in pain **sometimes**	You are in pain **sometimes**
Tiredness	You **never** feel tired	You **never** feel tired
Worry	You feel worried **all of the time**	You feel worried **all of the time**
Participation	You can take part in the things you want to **sometimes**	You can **never** take part in the things you want to
Feeling good about yourself	You **never** feel good about yourself	You **never** feel good about yourself
Which do you prefer?	○ Life A	○ Life B

where adjacent inconsistent coefficients were merged into a single variable. The use of the consistent model ensures a value set where a worsening in health leads to the same or lower utility values and is a widely used approach (for example[19,20,37,45,46]).

Robustness of the results was examined by estimating the models excluding respondents who may not have understood or engaged with the survey: respondents who did not choose the dominant profile in the practice question; respondents who did not choose the dominant profile in the dominance question; respondents who found the survey difficult to understand; and respondents who found the survey difficult to answer.

Preference heterogeneity, where preferences vary across respondents, was examined using interaction effects for sex, age, health, and employment status, and their sign, significance, and impact on anchored coefficients was considered.

Results

Classification system

Eight items from the DMD-QoL were selected for the HSC (with duration forming the maximum, ninth attribute). The data used to help aid item selection is included in Table 1. Selection was conducted for each factor (physical function, social participation, psychological impact) separately.

For physical functioning, out of 3 possible items, the items "I found it hard to use my hands" (difficulty using hands) and "I found it hard to breathe" (difficulty breathing) were selected for the HSC and "I found it hard to eat" (difficulty eating) was dropped. The former 2 items produced the largest spread across the latent factor and difficulty eating correlated the most highly with the other items (particularly use of hands, $\rho = .75$). Although all items in the physical functioning subscale displayed disordered thresholds, they did not if analyzed in combination with the other DMD-QoL items as an overall scale.[14] Furthermore, it was considered more important to retain representation of this factor in the HSC than to omit all physical functioning items based on this criterion.

For social participation, out of 3 possible items, the items "I found it hard to get around" (mobility) and "I could take part in the things I wanted to" (participation) were taken forward for the HSC and item "I could take part in things with my friends" (friends) was dropped. Friends had inadequate item fit, the highest average intra-factor correlation, and also had the highest proportion of missing data (5.88%). Mobility was considered a critical item in DMD-QoL development work.[14] Although mobility and friends had the largest spread across the latent continuum, the difference was marginal. Furthermore, participation was considered a broader term than friends and potentially inclusive of it, so was preferred for inclusion in the HSC (participation and friends correlated at $\rho = 0.59$).

Finally, for psychological impact, 4 of 8 possible items were selected for the HSC: "I was in pain" (pain), "I felt tired" (tiredness), "I felt good about myself" (feeling good about yourself), "I felt worried" (worry). The items "I felt unhappy" (unhappiness) and "I found it hard to talk to people" (communication) were eliminated as they had inadequate item fit and the latter was noted as potentially ambiguous in earlier development work.[14] The items "I felt embarrassed" (embarrassment) and "I felt angry" (anger) were eliminated as it was perceived to be too situational, with potential implausibility in feeling embarrassed or angry "all of the time" (reflected in the lowest percentage of responses at the floor for these items). Both worry and feeling good

Table 1. Summary of psychometric and IRT data used to help item selection.

Item	Psychometric analysis				IRT analysis (rating scale model)					
	% at floor (all of the time)	% at ceiling (never)	% missing data	Average intra-factor correlation*	Item location	Item range on the latent dimension	Infit MSQ	Outfit MSQ	Disordered thresholds	DIF[†]
Physical functioning										
I found it hard to use my hands	6.86	41.18	0.00	0.68	1.79	0.59-3.73	0.59	0.59	Yes	
I found it hard to eat	6.86	57.84	2.94	0.69	1.02	−0.17 to 2.96	0.77	0.66	Yes	
I found it hard to breathe	1.96	81.37	0.98	0.62	-0.60	−1.80 to 1.34	1.21	0.96	Yes	
Psychological impact										
I was in pain	1.96	37.25	0.00	0.32	1.64	−0.69 to 4.34	1.19	1.30		
I felt tired	5.88	3.92	0.00	0.28	3.42	1.09 to 6.12	0.86	0.85		Age
I found it hard to talk to people	5.88	43.14	0.00	0.34	1.70	−0.63 to 4.40	1.54	1.39		
I felt good about myself	15.69	4.90	2.94	0.38	3.12	0.79 to 5.82	0.93	0.96		
I felt unhappy	1.96	18.63	0.00	0.44	2.12	−0.22 to 4.82	0.49	0.47		
I felt embarrassed	0.98	36.27	2.94	0.37	1.60	−0.73 to 4.30	0.88	0.89		
I felt worried	5.88	14.71	0.98	0.44	2.78	0.45 to 5.48	0.64	0.61		
I felt angry	0.98	24.51	0.00	0.43	2.27	−0.06 to 4.97	0.69	0.68		
Social participation										
I found it hard to get around	19.61	19.61	0.00	0.51	1.96	−0.41 to 3.97	0.91	0.94		
I could take part in the things I wanted to	16.67	5.88	1.96	0.51	2.26	−0.12 to 4.27	0.72	0.73		
I could take part in things with my friends	10.78	8.82	5.88	0.60	2.90	0.53-4.91	0.42	0.43		

Note. N = 102.

DIF indicates differential item functioning; IRT, item response theory; MSQ, mean squared residual.

*Based on average correlation with other items in the same factor.

[†] DIF tested by report method (self-report vs proxy) and age (≥16 y vs <16 y). DIF not assessed for the physical and social subscales due to having <4 items.

about yourself were considered important to include given their prominence in earlier qualitative work.[14] Tiredness and pain were selected to enhance spread across the latent factor and both items showed good independence, motivating their inclusion in the HSC. Although tired was the only item to show potential DIF by age, with younger people reporting more tiredness, there are potential explanations for this result, including that children are more mobile and may experience greater fatigue as a consequence.[14] Accordingly, it was considered important to retain the item in the HSC due to its spread in the IRT analysis. The classification system is detailed in Fig. 2.

The sample

The socioeconomic and health characteristics of the sample are presented and compared with the UK general population in Table 2. The sample has 1043 respondents and is representative of the UK population in terms of age and sex. Approximately half of the sample has a university degree or equivalent qualification. Compared to the UK population, the sample has smaller proportions of respondents who are employed, retired or students, and has a proportion of respondents who are furloughed (because the survey was undertaken during the coronavirus 2019 [COVID-19] pandemic). The majority of respondents are either in very good or good health, with mean EQ-5D-5L utility of 0.792 (scored using[47]). DMD-QoL responses demonstrate the health problems present in the sample, where many respondents reported pain, tiredness, worry, and problems with both participation and feeling good about themselves. The majority of participants completed the survey on a laptop or desktop (93.5%) with the remainder completing the survey using a tablet (6.5%).

Understanding and engagement

The proportion of respondents reporting that the DCE tasks were difficult to understand was relatively small (12.7%), whereas the proportion of respondents reporting that they found the DCE tasks hard to answer was large (39.6%). This indicates that while most respondents understood the tasks, a proportion of respondents found it difficult to choose which health profile they thought was best. The majority of respondents correctly chose the dominant health profile in the practice DCE task (90.1%) and dominance DCE task embedded in the survey (86.6%).

Regression analysis

The regression analyses are reported in Table 3. Results are reported for the standard model that estimates coefficients for all interactions between the severity levels 2, 3, and 4 for every dimension with duration, and also for a fully consistent model where adjacent inconsistent coefficients are merged into a single variable.

Figure 2. DMD-QoL-8D classification system for valuation. Participation and feeling good about yourself are both reverse ordered. The ordering of some of the items has been changed in comparison to the order in which the items appear in the DMD-QoL. Labels in brackets and italics show the labels used in Table 3 and Figure 3 to refer to these dimensions.

Dimension	Level	Description
Mobility	1	You **never** find it hard to get around
	2	You find it hard to get around **sometimes**
	3	You find it hard to get around **a lot of the time**
	4	You find it hard to get around **all of the time**
Difficulty using hands *(hands)*	1	You **never** find it hard to use your hands
	2	You find it hard to use your hands **sometimes**
	3	You find it hard to use your hands **a lot of the time**
	4	You find it hard to use your hands **all of the time**
Difficulty breathing *(breathe)*	1	You **never** find it hard to breathe
	2	You find it hard to breathe **sometimes**
	3	You find it hard to breathe **a lot of the time**
	4	You find it hard to breathe **all of the time**
Pain	1	You are **never** in pain
	2	You are in pain **sometimes**
	3	You are in pain **a lot of the time**
	4	You are in pain **all of the time**
Tiredness *(tired)*	1	You **never** feel tired
	2	You feel tired **sometimes**
	3	You feel tired **a lot of the time**
	4	You feel tired **all of the time**
Worry *(worrird)*	1	You **never** feel worried
	2	You feel worried **sometimes**
	3	You feel worried **a lot of the time**
	4	You feel worried **all of the time**
Participation *(take part)*	1	You feel take part in the things you want to **all of the time**
	2	You can take part in the things you want to **a lot of the time**
	3	You can take part in the things you want to **sometimes**
	4	You can **never** take part in the things you want to
Feeling good about yourself *(feel good)*	1	You feel good about yourself **all of the time**
	2	You feel good about yourself **a lot of the time**
	3	You feel good about yourself **sometimes**
	4	You **never** feel good about yourself

For the latent estimates (unanchored models) in the standard model, the coefficients for all levels of all dimensions multiplied by duration are negative as expected, with the exception of participation level 2 and feel good about yourself levels 2 and 3. All other coefficients are logically consistent, whereas severity worsens, the utility decrement increases. The duration coefficient (LY) has the expected positive coefficient, demonstrating that individuals prefer to live longer meaning that higher levels of duration have higher utility. The consistent model merges participation levels 1 and 2 as well as feeling good about yourself levels 1 and 2. This means that for these dimensions there is no difference in utility between the levels all of the time and a lot of the time. The consistent model is logically consistent across all dimensions and 18 of 25 coefficients are significant. The inconsistent coefficients are for the milder levels of the dimensions and in particular for the tired dimension.

The anchored model is reported for the consistent model, where the utility decrements are anchored onto the 1 to 0 full health-dead scale. These are also plotted on Fig. 3. Across all the dimensions, pain and feeling good about yourself have the largest utility decrements for level 4, whereas the tired dimension has very small utility decrements for all severity levels (that are noticeably smaller than all other dimensions), meaning that these are the most and least important dimensions respectively in terms of their impact on utility.

Utility values for each health state are generated by adding the sum of the utility decrements to 1. For example, health state 33211111 has a utility value of 0.702 (calculated as 1+(−0.209−0.084−0.005)). The best state defined by the classification system has a value of 1 and the worst state defined by the classification system has a value of −0.559 (worse than dead).

Robustness analyses are reported in Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.03.007, where the standard regression model is estimated excluding respondents are who may not have understood or engaged with the DCE tasks. These models have the same logical inconsistencies and the same insignificant coefficients as the standard model estimated using the full sample, suggesting that the models are robust to potential issues of participant engagement or comprehension.

Table 2. Sample of DCE survey respondents.

Sociodemographic characteristics Rasch measurement transactions	UK general population, %[†]	Sample (n = 1043), %
Sex		
Male	49.1	48.7
Female	50.9	51.1
Other/nonbinary		0.2
Age, y		
Mean age (SD)		47.9 (17.6)
Age 18-44	46.6[*]	43.6
45-64	32.5	34.0
65+	20.9	22.3
Highest education/school qualification		
PhD or equivalent doctoral level qualification		3.8
Masters or equivalent higher degree level qualification		14.8
Bachelors or equivalent first degree level qualification		33.0
A-level or equivalent post-secondary level qualification		24.2
GCSE or equivalent secondary level qualification.		19.9
None of the above/prefer not to say		4.3
Employment status		
Employed	61.7	57.0
Retired	13.9	22.1
Looking after home or family	4.3	3.9
Student	9.3	4.3
Seeking work		1.3
Unemployed	4.4	3.2
Furloughed[‡]		3.7
Long term sick	4.3	3.6
None of the above/prefer not to say	2.2	0.9
Health		Sample (n = 1043), %
General health	Excellent	11.1
	Very good	33.0
	Good	34.2
	Fair	16.2
	Poor	5.3
	Prefer not to say	0.2
EQ-5D-5L	Mean (SD)	0.792 (0.236)
Understanding and engagement		
Difficulty to answer DCE tasks	Very difficult to answer	5.5
	Quite difficult to answer	34.1
	Neither difficult nor easy to answer	19.0
	Fairly easy to answer	28.1
	Very easy to answer	13.3
Difficulty to understand DCE tasks	Very difficult to understand	2.7
	Quite difficult to understand	10.0
	Neither difficult nor easy to understand	17.6
	Fairly easy to understand	35.1
	Very easy to understand	34.6
Time in minutes taken to complete survey	Mean (SD)	9.61 (6.31)
	Interquartile range	5.55 to 11.95
	Median	8.14
Practice question	Selected dominant option	90.1%
Dominance question	Selected dominant option	86.6%

DCE indicates discrete choice experiment; GCSE, General Certificate of Secondary Education.

[*] Age distribution is here reported as the percentage of all adults aged ≥18 years old.

[†] Statistics for England in the Census 2011. The census includes persons ≥16 years old, whereas this study only surveys persons aged ≥18 years old.

[‡] The survey was conducted July to August 2020 when in the United Kingdom a furlough scheme was in operation where due to the coronavirus 2019 pandemic some employees were placed on temporary leave and the UK government paid 80% of their wages.

Table 3. Regression analysis of DCE survey responses.

	Standard model	Consistent model		Anchored coefficients of consistent model
Mobility2_LY	−0.025*	−0.026*	Mobility2	−0.064*
	(0.008)	(0.007)		(0.023)
Mobility3_LY	−0.084*	−0.084*	Mobility3	−0.209*
	(0.000)	(0.000)		(0.021)
Mobility4_LY	−0.089*	−0.085*	Mobility4	−0.213*
	(0.000)	(0.000)		(0.023)
Hands2_LY	−0.021†	−0.020†	Hands2	−0.051†
	(0.018)	(0.023)		(0.022)
Hands3_LY	−0.033*	−0.034*	Hands3	−0.084*
	(0.000)	(0.000)		(0.022)
Hands4_LY	−0.076*	−0.077*	Hands4	−0.191*
	(0.000)	(0.000)		(0.022)
Breathe2_LY	−0.004	−0.002	Breathe2	−0.005
	(0.612)	(0.831)		(0.022)
Breathe3_LY	−0.040*	−0.038*	Breathe3	−0.095*
	(0.000)	(0.000)		(0.022)
Breathe4_LY	−0.078*	−0.075*	Breathe4	−0.186*
	(0.000)	(0.000)		(0.022)
Pain2_LY	−0.044*	−0.046*	Pain2	−0.114*
	(0.000)	(0.000)		(0.023)
Pain3_LY	−0.099*	−0.096*	Pain3	−0.240*
	(0.000)	(0.000)		(0.021)
Pain4_LY	−0.115*	−0.116*	Pain4	−0.288*
	(0.000)	(0.000)		(0.025)
Tired2_LY	−0.001	−0.001	Tired2	−0.003
	(0.938)	(0.927)		(0.029)
Tired3_LY	−0.018‡	−0.014	Tired3	−0.035
	(0.091)	(0.170)		(0.025)
Tired4_LY	−0.022†	−0.018†	Tired4	−0.045†
	(0.016)	(0.047)		(0.022)
Worried2_LY	−0.007	−0.006	Worried2	−0.015
	(0.430)	(0.479)		(0.020)
Worried3_LY	−0.024†	−0.023†	Worried3	−0.057†
	(0.011)	(0.015)		(0.023)
Worried4_LY	−0.066*	−0.066*	Worried4	−0.164*
	(0.000)	(0.000)		(0.028)
Take_Part2_LY	0.018*		Take_Part2	
	(0.074)			
Take_Part3_LY	−0.011	−0.020†	Take_Part3	−0.051†
	(0.331)	(0.040)		(0.024)
Take_Part4_LY	−0.074*	−0.083*	Take_Part4	−0.206*
	(0.000)	(0.000)		(0.018)
Feel_Good2_LY	0.028*		Feel_Good2	
	(0.004)			
Feel_Good3_LY	0.000	−0.014	Feel_Good3	−0.034
	(0.980)	(0.110)		(0.021)

continued on next page

Table 3. Continued

	Standard model	Consistent model		Anchored coefficients of consistent model
Feel_Good4_LY	−0.092*	−0.107*	Feel_Good4	−0.266*
	(0.000)	(0.000)		(0.020)
LY	0.384*	0.401*		
	(0.000)	(0.000)		
Observations	18 770	18 770		
Log likelihood	−5249	-5255		
Rho-squared	0.193	0.192		
Significant coefficients	20/25	18/23		

Note. For the standard model and consistent model *P* values are in parentheses. For the anchored coefficients standard errors are in parentheses.
Breathe indicates difficulty breathing; DCE, discrete choice experiment; hands, difficulty using hands; feel good, feeling good about yourself; LY, life years or duration; take part, participation; tired, tiredness; worried, worry.
*$P < .01$.
†$P < .05$.
‡$P < .1$.

Models examining preference heterogeneity across sociodemographic and health characteristics are reported in the Appendix Tables 2 and 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.03.007. In general, males, younger participants aged 18 to 24 years and employed participants have smaller anchored utility decrements, while older participants aged ≥65 years old and participants in fair or poor general health have larger utility decrements, though many coefficients were insignificant.

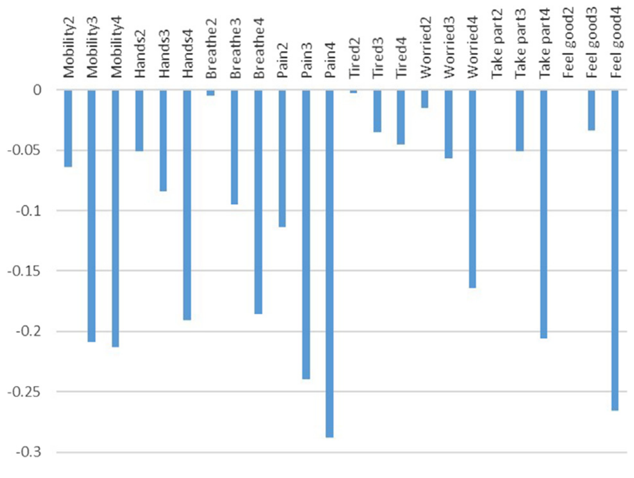
Comparing DMD-QoL-8D and EQ-5D-5L

The consistent model was used to generate DMD-QoL-8D utilities on the data used to derive the classification system.[14] For the 57 participants with both DMD-QoL-8D and EQ-5D-5L data, DMD-QoL-8D had mean (SD) 0.675(0.256) with interquartile range (IQR) 0.304, and EQ-5D-5L had mean (SD) 0.206(0.346) and IQR 0.643. The measures are strongly correlated at 0.65.

Discussion

This report has presented DMD-QoL-8D, a DMD-specific PBM. The research has derived the health state classification system from the DMD-QoL, valued health states from the DMD-QoL using DCE with a duration attribute, DCE_{TTO}, and modeled these results to generate utility values for all health states defined by the classification system. The generation of this PBM has the advantage that it is based upon a measure that has been developed with input from patients, caregivers and clinicians and is patient-focused and relevant for people with DMD. The measure can be used to estimate utility values to calculate QALYs for assessing the cost-effectiveness of new and existing interventions in DMD. The classification system was valued by a representative sample of the UK general population, as recommended to inform decision making for agencies such as NICE.[17]

Figure 3. Plot of anchored coefficients to generate value set.



The consistent model is recommended to generate utility values for the DMD-QoL-8D. The model has a large number of significant and consistent coefficients, noting that there are 2 merged variables for participation levels 1 and 2 and feeling good about yourself levels 1 and 2. The model includes all responses, and does not exclude any participants on the basis of engagement and understanding or impose any potentially subjective judgements around whose preferences are appropriate or inappropriate, and benefits from a large sample size of >1000 participants. The utility values have a large range from 1 to −0.559, where the worst state is considerably lower than dead at zero. This range in utility values is similar to UK values for EQ-5D-3L (1 to −0.594).[48]

Overall, the sample had good understanding, with the majority of respondents reporting that the DCE tasks were easy to understand (69.7%), and the majority of respondents correctly choosing the dominant option in the practice question (90.1%) and purposively designed dominance question (86.6%). This is higher than the mean of 82% found in a recent article that assessed dominance questions across 22 studies whilst developing and testing a tool to assess internal validity of DCE data.[49] Many respondents did state that the DCE tasks were difficult to choose an answer (39.6%), which may be expected since this requires the imagination of hypothetical health states, trading between quality and quantity of life and the simultaneous consideration of many pieces of information.

Two of the dimensions that capture QoL (participation and feeling good about yourself) rather than health have amongst the largest utility decrements for the level 4, but for level 2 this was merged with level 1 meaning that there is no distinction in utility between all of the time and a lot of the time for these dimensions. There is a possibility that this may have been affected by the reverse ordering of these 2 dimensions because all of the time is the best level for these 2 dimensions but the worst level for all of the other dimensions. However, the coefficients for levels 3 and 4 are logically consistent as expected, and significant with the exception of feeling good about yourself level 3. Tired has smaller coefficients across all severity levels in comparison to all other dimensions. Tiredness or vitality has also been found to have small utility decrements in other PBM, including recently valued measures in the UK of generic SF-6Dv2[38] and EORTC-QLU-C10D for cancer.[36]

The factor analysis of the DMD-QoL identified 3 factors rather than the 7 themes (physical aspects, social relationships, autonomy, identity, daily activities, feelings and emotions, healthcare and support) identified in the qualitative analysis undertaken with people with DMD that was used to generate DMD-QoL items. The selected items from the DMD-QoL that form the DMD-QoL-8D classification system cover 5 of the 7 themes, and the 2 themes of daily activities and social relationships are not directly represented, but items are included, which otherwise relate to them. Daily activities was captured in the DMD-QoL by a single item around difficulty eating, and this correlated highly with difficulty using hands and difficulty breathing, which were both included in the classification system. Social relationships were captured using items of difficulty talking to people and being able to take part in things with friends. The former was noted as potentially ambiguous in earlier development work[14] and the latter was not independent to other items in its factor (that were included in the classification system), had the highest proportion of missing data and correlated highly with other items that were included in the classification system. Therefore, given the need for item reduction in developing the PBM, these items were not considered as favorable for inclusion. Although this means that the classification system does not include items purposively worded to capture these 2 themes as coded in the qualitative work, the classification system does include items that are correlated with the absent themes according to the psychometric analyses conducted here.

The data used for the psychometric analyses that informed the derivation of the classification system has a relatively small sample (n = 102) and is the only currently available DMD-QoL dataset. Although this is not unexpected, because there are significant challenges in the collection of QoL data in rare conditions, it does mean that the classification system has not been validated in an independent dataset. Therefore, when new DMD-QoL data becomes available it is recommended that the classification system and factor structure of the DMD-QoL are validated. This data were also used to compare EQ-5D-5L and DMD-QoL-8D utilities, where EQ-5D-5L has lower mean and larger IQR. However, this is based on a small sample (n = 57). Comparison of the psychometric performance of DMD-QoL-8D and EQ-5D-5L using an independent dataset with larger sample size, preferably associated with an intervention to be able to compare responsiveness, is recommended. However, the findings here suggest that DMD-QoL-8D and EQ-5D-5L produce different utility values and hence the mixing of utility values from both measures in an economic model would need to be carefully managed, because both the utility and change in utility is expected to vary widely across the measures. Where comparability across conditions and interventions is required, one option is to use a generic preference-based measure in base case economic analyses to enable that comparability, and DMD-QoL-8D in sensitivity analyses because it has greater content validity and includes aspects deemed important to patients.

DMD affects boys and adults, and hence the DMD-QoL has been designed for use for all people with DMD regardless of their age. There is a growing literature around the valuation of PBM for children and adolescents, with no consensus as to how they should be valued,[50] but this raises questions about whose values (eg, adults or adolescents) and from which perspective (for themselves or imagining, say, a 10-year-old child). This measure was valued using a representative sample of adults from their own perspective, because this has the advantage that it is both what is done for adult preference-based measures and what is recommended for adult measures by agencies such as NICE.[17] Adults valuing from their own perspective also has been used to value generic HUI3, a measure that can be used both for children and adults,[51] although note a different approach was used for HUI2 (which is also a measure that can be used for children and adults).[52] However, one option is that the measure could have been valued from a sample including both older adolescents and adults, since these are the population (as well as children) that have DMD.

The DCE survey was conducted during the COVID-19 pandemic, and it is possible that this may have impacted on preferences. In particular, some of the attributes were impacted by COVID-19, including difficulty breathing, and also on the wider population as a result of lockdown, including participation and potentially worry. The impact of the pandemic on health preferences is currently unknown, although due to the nature of the global pandemic it is likely that the impact on preferences will be experienced for some time. One possibility is to examine whether equivalent preferences are found in a future survey where the same questions are answered, potentially on a smaller sample. It is not expected that the mode of administration will have impacted on preferences, because online surveys have been commonly used in recent years and it is expected that they will become increasingly popular in the years during and after the COVID-19 pandemic. There is an additional possibility that participants undertaking the survey differed across their unmeasurable characteristics to those who typically complete online surveys or complete interviews in their own home. During the pandemic everyone was encouraged to stay at home, and this may have positively impacted on people's availability and willingness to complete online surveys particularly in the area of health.

Models exploring preference heterogeneity indicated significant differences for age, sex, employment status, and general health characteristics. However, the selected model does not include interaction effects to reflect these since the model is intended to reflect the preferences of the general population using a single value set. Participants were purposively sampled to be representative for age and sex and the sample is representative for the proportion of employed individuals.

Potential limitations of the study include those arising through the use of an online sample recruited via an existing online panel managed by a market research agency. Although the use of these samples is becomingly increasing common in health preference research, concerns have been raised around the true representativeness of the sample that will not include the computer illiterate or those with no internet access. Our sample has a high proportion (51.6%) of people with a degree, suggesting the sample is not fully representative regarding education. However, importantly the online administration would include people that are shielding due to COVID-19, which would have been missed using other modes of administration. Second, lack of attendance to all dimensions and duration in the DCE task is a criticism of the $\mathrm{DCE}_{\mathrm{TTO}}$ technique and may be further emphasized when the dimensions that vary across the 2 profiles are highlighted in yellow thus encouraging respondents to focus on the levels that change rather than all dimensions included in the profiles. However, the use of highlighting is a recognized technique when there are a large number of attributes in the DCE task. One final limitation is that the approach adopted here differs to the elicitation technique used for EQ-5D-3L or EQ-5D-5L, which is the recommended measure for agencies such as NICE.[17] However, the same technique of DCE_{TTO} has recently been used to value other measures in the UK including generic SF-6Dv2[38] and EORTC-QLU-C10D in cancer,[36] and it is emerging as a widely used technique to value PBM more generally.[31,32]

This study has generated preference weights for the DMD-QoL-8D, a DMD-specific PBM that has been derived to generate utility values for people with DMD. These results can be used to inform cost-effectiveness analyses of interventions in DMD.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2021.03.007.

Article and Author Information

Accepted for Publication: March 9, 2021

doi: https://doi.org/10.1016/j.jval.2021.03.007

Author Affiliations: School of Health and Related Research (ScHARR), University of Sheffield, England, UK (Rowen, Powell, Mukuria, Carlton, Brazier); School of Public Health, Curtin University, Perth, Australia (Norman)

Correspondence: Donna Rowen, PhD, School of Health and Related Research (ScHARR), University of Sheffield Regent Court, 30 Regent St, Sheffield, South Yorkshire S1 4DA, United Kingdom. Email: d.rowen@sheffield.ac.uk

Author Contributions: *Concept and design:* Rowen, Powell, Mukuria, Carlton, Norman, Brazier
Acquisition of data: Rowen, Powell, Mukuria, Carlton, Norman, Brazier
Analysis and interpretation of data: Rowen, Powell, Mukuria, Carlton, Norman, Brazier
Drafting of the manuscript: Rowen, Powell, Mukuria, Carlton, Norman, Brazier
Critical revision of the paper for important intellectual content: Rowen, Powell, Mukuria, Carlton, Norman, Brazier
Statistical analysis: Rowen, Powell, Mukuria, Carlton, Norman, Brazier

Conflict of Interest Disclosures: Drs Rowen, Powell, Mukuria, Carlton, and Brazier reported receiving grants from the Project HEalth Research Collaboration United in Leading Evidence Synthesis (HERCULES) consortium during the conduct of the study. Dr Norman is an editor for *Value in Health* and had no role in the peer review process of this article. No other disclosures were reported.

Funding/Support: This study was funded by Project HERCULES consortium. The Project HERCULES consortium is funded by Duchenne UK, Pfizer, PTC Therapeutics, Roche, Summit Therapeutics plc, Sarepta Therapeutics Inc, Wave Lifesciences USA Inc, Solid Biosciences, Catabasis Pharmaceuticals, and Santhera Therapeutics.

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

Acknowledgment: We would like to thank all participants who took part in this research, both those with Duchenne and from the general population. We would also like to acknowledge everyone who took part in our Patient and Public Involvement and Engagement advisory groups and patient advocates within the Duchenne community, including the DMD Pathfinders and Muscular Dystrophy UK, who helped facilitate recruitment. We would like to thank everyone at Duchenne UK and the HERCULES consortium for their considerable support and assistance including Emily Crossley, Fleur Chandler, Josie Godfrey, and Dala Jenkin. We would also like to thank the 5 National Health Service Trusts and affiliated clinicians that helped to recruit for the project and provided expertise and guidance, including Anne-Marie Childs, Michela Guglieri, Anirban Majumdar, Anna Mayhew, Stefan Spinty, Volker Straub and Ros Quinlivan. Our thanks also go to David Abbott who provided feedback on the project.

REFERENCES

1. Rowen DL, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International regulations and recommendations for utility data for health technology assessment. *PharmacoEcon*. 2017;35(Suppl 1):11–19.
2. Kennedy-Martin M, Slaap B, Herdman M, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ*. 2020;21(8):1245–1257.
3. Stark AE. Determinants of the incidence of Duchenne muscular dystrophy. *Ann Transl Med*. 2015;3(19):287.
4. Wong SH, McClaren BJ, Dalton Archibald A, et al. A mixed methods study of age at diagnosis and diagnostic odyssey for Duchenne muscular dystrophy. *Eur J Hum Genet*. 2015;23(10):1294–1300.
5. Landfeldt E, Thompson R, Sejersen T, McMillan HJ, Kirschner J, Lochmüller H. Life expectancy at birth in Duchenne muscular dystrophy: a systematic review and meta-analysis. *Eur J Epidemiol*. 2020;35:643–653.
6. Flores D, Ribate MP, Montolio M, Ramos FJ, Gómez M, García CB. Quantifying the economic impact of caregiving for Duchenne muscular dystrophy (DMD) in Spain. *Eur J Health Econ*. 2020;21:1015–1023.
7. Uttley L, Carlton J, Buckley Woods H, Brazier J. A review of quality of life themes in Duchenne muscular dystrophy for patients and carers. *Health Qual Life Outcomes*. 2018;16(1):237.
8. Uzark K, King E, Cripe L, et al. Health-related quality of life in children and adolescents with Duchenne muscular dystrophy. *Pediatrics*. 2012;130(6):e1559–e1566.
9. Lue YJ, Chen SS, Lu YM. Quality of life of patients with Duchenne muscular dystrophy: from adolescence to young men. *Disabil Rehabil*. 2017;39(14):1408–1413.
10. Bann CM, Abresch RT, Biesecker B, et al. Measuring quality of life in muscular dystrophy. *Neurology*. 2015;84(10):1034–1042.
11. Straub VE, Mercuri E, DMD outcome measure study group. Report on the workshop: meaningful outcome measures for Duchenne muscular dystrophy. *Neuromuscul Disord*. 2018;28(8):690–701.
12. Mokkink L, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcome. *J Clin Epidemiol*. 2010;63(7):737–745.
13. Powell PA, Carlton J, Woods HB, Mazzone P. Measuring quality of life in Duchenne muscular dystrophy: a systematic review of the content and structural validity of commonly used instruments. *Health Qual Life Outcomes*. 2020;18(1):1–26.
14. Powell PA, Carlton J, Rowen D, Chandler F, Guglieri M, Brazier J. Development of a new quality of life measure for Duchenne muscular dystrophy using mixed methods: the DMD-QoL. *Neurology*. Online ahead of print.
15. Powell PA, Carlton J, Rowen D, Brazier J. Producing a preference-based quality of life measure for people with Duchenne muscular dystrophy: a mixed-methods study protocol. *BMJ Open*. 2019;9(3):e023685.
16. Szabo SM, Audhya IF, Malone DC, Feeny D, Gooch KL. Characterizing health state utilities associated with Duchenne muscular dystrophy: a systematic review. *Qual Life Res*. 2020;29(3):593–605.
17. NICE (National Institute for Health and Care Excellence). *Guide to the Methods Of Technology Appraisal*. London: NICE; 2013.
18. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev*. 1956;63(2):81–97.
19. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a preference-based measure for myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015;18(6):846–855.
20. Rowen D, Brazier J, Young T, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value Health*. 2011;14(5):721–731.
21. Rowen DL, Labeit A, Stevens K, et al. Estimating a preference-based single index measuring the quality of life impact of self-management for diabetes. *Med Decis Making*. 2018;38(6):699–707.
22. Brazier JE, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess*. 2012;16(32):1–114.
23. Brazier J, Mulhern B, Bjorner JB, et al. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. *Med Care*. 2020;58(6):557–565.
24. Wright BD, Linacre JM, Gustafson JE, Martin-Lof P. Reasonable mean-square fit values. *Rasch Measurement Transactions*. 1994;8(3):370.
25. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press; 2016.
26. Choi SW, Gibbons LE, Crane PK. Lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *J Stat Softw*. 2011;39(8):1–30.

27. R Development Core Team. *R: A language and environment for statistical computing*. R Development Core Team; 2013.
28. Revelle W. *Psych: procedures for personality and psychological research*. Evanston: Northwestern University; 2018.
29. Rosseel Y. lavaan: an R package for structural equation modeling. *J Stat Softw*. 2012;48:1–36.
30. Mair P, Hatzinger R. Extended Rasch modeling: the eRm package for the application of IRT models in R. *J Stat Softw*. 2007;20.
31. Bahrampour M, Byrnes J, Norman R, Scuffham PA, Downes M. Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods. *Eur J Health Econ*. 2020;21:983–992.
32. Mulhern B, Norman R, Street D, Viney R. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. *PharmacoEcon*. 2019;37(1):29–43.
33. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *J Health Econ*. 2012;31(1):306–318.
34. Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. *Appl Health Econ Health Policy*. 2013;11(3):287–298.
35. Norman R, Viney R, Brazier J, et al. Valuing SF-6D health states using a discrete choice experiment. *Med Decis Making*. 2014;34(6):773–786.
36. Norman R, Mercieca-Bebber R, Rowen D, et al. UK utility weights for the EORTC QLU-C10D. *Health Econ*. 2019;28(12):1385–1401.
37. Rowen DL, Mulhern B, Stevens K, Vermaire E. Estimating a Dutch value set for the paediatric preference-based CHU-9D using a discrete choice experiment with duration. *Value Health*. 2018;21(10):1234–1242.
38. Mulhern BJ, Bansback N, Norman R, Brazier J. Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. *Med Care*. 2020;58(6):566–573.
39. Norman R, Viney R, Aaronson NK, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. *Qual Life Res*. 2016;25(3):637–649.
40. Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Effect of level overlap and color coding on attribute nonattendance in discrete choice experiments. *Value Health*. 2018;21(7):767–771.
41. Choice Metrics. *Ngene software for experimental design*. Sydney: Choice Metrics; 2018.
42. Cook RD, Nachtrheim CJ. A comparison of algorithms for constructing exact D-optimal designs. *Technometrics*. 1980;22(3):315–324.
43. Orme B. *Sample size issues for conjoint analysis studies. Sawthooth software research paper series*. 10. Squim, WA, USA: Sawthooth Software Inc; 1998.
44. Payne K, Fargher EA, Roberts SA, et al. Valuing pharmacogenetic testing services: a comparison of patients' and healthcare professionals' preferences. *Value Health*. 2011;14:121–134.
45. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851–859.
46. Stevens K. Valuation of the Child Health Utility 9D index. *PharmacoEcon*. 2012;30(8):729–747.
47. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15(5):708–715.
48. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
49. Johnson FR, Yang J-C, Reed SD. The internal validity of discrete choice experiment data: a testing tool for quantitative assessments. *Value Health*. 2019;22(2):157–160.
50. Rowen D, Rivero-Arias O, Devlin N, Ratcliffe J. Review of valuation methods of preference-based measures of health for economic evaluation in child and adolescent populations: where are we now and where are we going? *PharmacoEcon*. 2020;38:325–340.
51. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark3 system. *Med Care*. 2002;40(2):113–128.
52. Torrance G, Feeny D, Furlong W, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2. *Med Care*. 1996;34:702–722.