

# Development of a New Quality of Life Measure for Duchenne Muscular Dystrophy Using Mixed Methods

## The DMD-QoL

Philip A. Powell, PhD, Jill Carlton, PhD, Donna Rowen, PhD, Fleur Chandler, MSc, Michela Guglieri, MD, and John E. Brazier, PhD

*Neurology*® 2021;96:e2438-e2450. doi:10.1212/WNL.0000000000011896

### Correspondence

Dr. Powell  
p.a.powell@sheffield.ac.uk

## Abstract

### Objective

Based on concerns about existing patient-reported outcome measures (PROMs) for assessing quality of life (QoL) in Duchenne muscular dystrophy (DMD), we describe the mixed methods development of a new QoL PROM for use in boys and men with DMD: the DMD-QoL.

### Methods

The DMD-QoL was developed in 3 stages. First, draft items were generated from 18 semi-structured qualitative interviews with boys and men with DMD, analyzed using framework analysis. Second, cognitive debriefing interviews with patients (n = 10), clinicians (n = 8), and patients' parents (n = 10) were undertaken, and a reduced item set was selected and refined. Third, psychometric data on the draft items from a cross-sectional online survey (n = 102) and stakeholder input from patients and patients' parents were used to produce the final questionnaire. Patient and public involvement and engagement was embedded throughout the process.

### Results

From an initial draft of 47 items, a revised set of 27 items was produced at stage 2, and this set was further refined at stage 3 to generate the DMD-QoL, a 14-item QoL PROM. The DMD-QoL is designed for use from 7 years of age by proxy report and from 10 years of age by self-report or proxy report. The final measure showed good psychometric properties.

### Conclusion

The DMD-QoL is a new 14-item QoL PROM for boys and men with DMD, with demonstrable content and face validity.

---

From the School of Health and Related Research (P.A.P., J.C., D.R., J.E.B.), University of Sheffield; Duchenne UK (F.C.), London; and John Walton Dystrophy Research Centre (M.G.), Institute of Genetic Medicine, University of Newcastle and Newcastle Hospitals NHS Foundation Trust, UK.

Go to [Neurology.org/N](https://www.neurology.org/N) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

The Article Processing Charge was funded by Duchenne UK under Project HERCULES.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

## Glossary

CFA = confirmatory factor analysis; CMQM = Comprehensive Model of QoL in Muscular Dystrophy; DIF = differential item functioning; DMD = Duchenne muscular dystrophy; EFA = exploratory factor analysis; GCS = Generic Core Scales; IRT = item response theory; MSA = Mokken scale analysis; NHS = National Health Service; PedsQL = Pediatric Quality of Life Inventory; PPIE = patient and public involvement and engagement; PROM = patient-reported outcome measure; QoL = quality of life; RMSEA = root mean square error of approximation.

Studies assessing health-related quality of life (QoL) in people with Duchenne muscular dystrophy (DMD) suggest that it is impaired compared to the general population.<sup>1</sup> However, the ability of existing measures to adequately capture QoL in DMD has been questioned,<sup>2</sup> with it posited that there is no optimal measure available.<sup>3</sup> A recent systematic review of the content and structural validity of QoL instruments used in DMD highlighted that no measure had high-quality evidence to support its use.<sup>4</sup>

Given the problems with existing measures, there is clear justification for a new QoL patient-reported outcome measure (PROM) for use in people with DMD, developed with demonstrable content validity. A further aim of the present work was to produce a QoL measure that could be subsequently adapted for use in the economic evaluation of DMD health interventions via cost utility analysis.<sup>5</sup> Specialized preference-based PROMs are required for this purpose. Existing, popular generic preference-based measures such as the EQ-5D have been found to have poor content validity in DMD,<sup>4</sup> suggesting that there are no existing tools with proven content validity that can be used to directly inform economic health care resource allocation decisions in DMD.

Our primary objective was to develop a new QoL PROM for use in boys and men with DMD ( $\geq 7$  years of age) and to assess its preliminary psychometric properties. The PROM has been developed rigorously, across a series of sequential stages, in accordance with best practice guidelines.<sup>6</sup> Patient and public involvement and engagement (PPIE) with patients and patients' parents has been embedded throughout (figure 1).

## Methods

The stages of development for the DMD-QoL are summarized in figure 1 and in a published protocol.<sup>5</sup> We used a mixed methods PROM development design involving 3 stages: qualitative interviewing (item generation), qualitative cognitive debriefing interviews (item selection), and a cross-sectional quantitative online survey (psychometric survey).

### Standard Protocol Approvals, Registrations, and Patient Consents

The research received ethics approval from the UK National Health Service (NHS; REC reference: 18/SW/0055). Informed consent was obtained from all participants (or guardians of participants) in the study.

### Stage 1: Item Generation

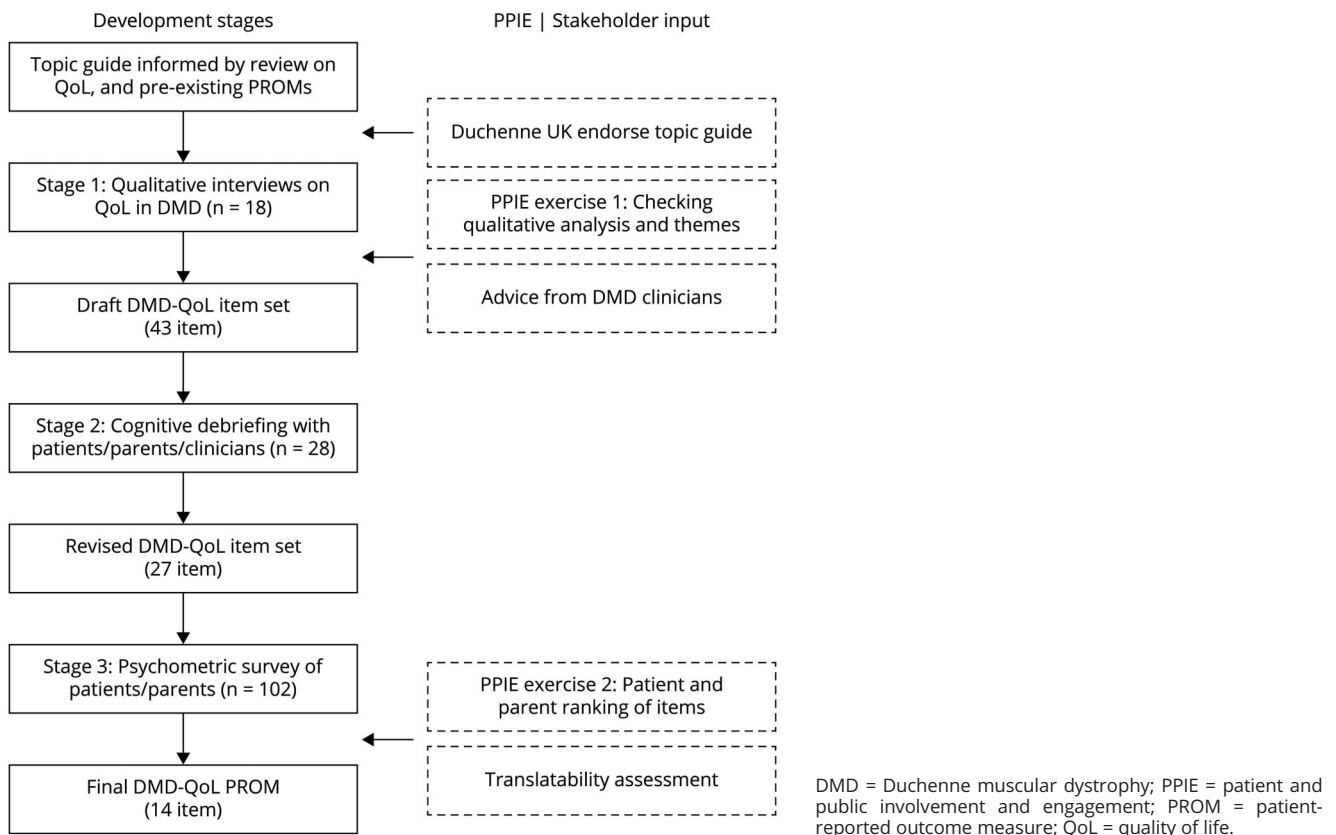
Items for the draft questionnaire were informed by semi-structured interviews with 18 boys and men with DMD from the United Kingdom between June 2018 and December 2018. Because we wanted to gain deeper insight into individuals' experiences of QoL, one-to-one interviews were preferred over focus groups. The sample was purposively recruited with a sampling grid (table 1) from 5 collaborating UK NHS sites (Alder Hey Children's NHS Foundation Trust, University Hospitals Bristol NHS Foundation Trust, Leeds Teaching Hospitals NHS Trust, Newcastle Upon Tyne Hospitals NHS Foundation Trust, University College London Hospitals NHS Foundation Trust), the charity Duchenne UK, and a patient support group (DMD Pathfinders). Eligibility criteria consisted of a confirmed (NHS Trusts) or self-disclosed (Duchenne UK/DMD Pathfinders) diagnosis of DMD. Multiple sources of recruitment were used to try to obtain as broad a participation as possible.

Recruitment continued until data saturation was reached (no new themes emerged in the interviews), which was sooner than anticipated in the study protocol<sup>5</sup> and encompassed a breadth of age and clinical characteristics (table 1). Participants were interviewed either face-to-face or online (via Skype) for a mean length of 53 minutes (SD 15 minutes) with a flexible approach adopted to help people participate in their preferred medium. A topic guide (available online: doi.org/10.15131/shef.data.12942998.v1, file A) was developed to cover aspects of QoL known to be potentially relevant to DMD<sup>1</sup> and was endorsed by PPIE representatives at Duchenne UK. As part of the interview, participants were shown examples of generic PROMs commonly used in cost-utility analyses (i.e., EQ-5D or EQ-5D-Y,<sup>7,8</sup> Health Utilities Index,<sup>9,10</sup> and/or Child Health Utility-9D<sup>11</sup>) and asked for their opinions on them, including their comprehensiveness.

Interviews were conducted by a chartered psychologist with experience in qualitative research and conducting research interviews. The interviewer was not known to the participants before the study. While the interviewer had knowledge of QoL questionnaires and their typical content, a conscious effort was made not to restrictively impose those assumptions on participants, facilitating their inductive contributions.

The interviews were audio recorded, transcribed verbatim, anonymized, and analyzed with a modified, iterative framework analysis to generate themes.<sup>12</sup> Framework analysis is a flexible approach, considered independent from any particular theoretical, philosophical, or epistemologic stance.<sup>12</sup> It was chosen

**Figure 1** Stages of Development of the DMD-QoL



because it enables a combination of deductive (from a priori QoL themes) and inductive (new themes from participants) analysis and is commonly used in qualitative health care research, including in projects developing PROMs from qualitative data.<sup>13,14</sup>

Coding was deductive (based on initial QoL themes covered in the topic guide) and inductive (emerging themes from the data) and was conducted on hard copy transcripts using the margins. Coding was split between the interviewer (who coded all transcripts) and another member

of the research team (who secondary coded 50% of the transcripts to enhance trustworthiness). Before coding, the researchers familiarized themselves with the data by reading and listening to the recordings of the transcripts being analyzed.

After 2 adult and 2 child transcripts had been secondary coded, the 2 researchers met to discuss their analysis and to derive an initial working thematic framework. This working framework was then applied to subsequent transcripts by the

**Table 1** Sampling Grid for the Stage 1 Qualitative Interviews

Lower limb function	Gross upper limb function	Respiratory function	Age group, y					
			7-9	10-14	15-19	20-29	30-39	≥40
✓	✓	✓	2	3	1	1		
x	✓	✓		3		1		
x	x	✓			2	1		
x	x	x				1	2	1
<b>Total</b>			2	6	3	4	2	1

Abbreviations: Lower limb function ✓/x = can/cannot walk independently without a mobility device; gross upper limb function ✓/x = can/cannot raise a hand to mouth to eat/drink independently; respiratory function ✓/x = can/cannot breathe without any ventilator support throughout the day. N = 18.

**Table 2** Summary of the Rules Used for Item Selection Across all Stages

Stage 1: item generation	Stage 2: initial item selection
1A. Reflects QoL in DMD	2A. Reapply rules from stage 1
At least 1 item per domain.	2B. Ensure comprehensibility
1B. Ease of completion	Remove/revise if not understood as intended.
Reading level appropriate ( $\geq 7$ y).	2C. Ensure relevance
Avoid double barrel (e.g., anxiety and depression).	Remove/revise if not considered relevant to DMD.
Avoid double negatives.	2D. Ensure comprehensiveness
Avoid jargon and colloquialisms (e.g., “felt blue”).	Consider participant suggestions for missing items.
Avoid making items too long.	Add item if it does not conflict with rules from Stage 1.
Avoid overly personal items (e.g., sexual activity).	Add item if it does not increase overlapping items.
Avoid items not relevant to all (e.g., work).	2E. Reduce redundancy/overlap
Avoid ethically inappropriate questions.	Choose preferred item from cognitive debriefing.
Avoid questions on external knowledge.	Retain similar items if people do distinguish them.
1C. Not value laden	<b>Stage 3: final item selection</b>
Avoid value-laden statements (i.e., “X is good”).	3A. Reapply rules from stage 1
Avoid items not predictive of good/bad QoL for all.	3B. Reflects QoL in DMD
1D. Interpersonal and intertemporal comparison	At least 1 item per original domain.
Avoid temporal comparisons (e.g., “as usual”).	At least 3 items per CMQM domain.
Avoid personal comparisons (e.g., “as my peers”).	3C. Psychometric performance
Avoid expectations (e.g., “things will get better”).	Good-fitting model (EFA/CFA).
Choose items focused on the current situation.	No cross-loadings (EFA/CFA).
Avoid trait items (e.g., “I have a bad temper”).	Minimize problems with local independence (MSA).
1E. Psychometric performance	Minimize problems with monotonicity (MSA).
Avoid items likely to suffer from DIF.	Minimize problems with IIO (MSA).
Items cover a range of severity within DMD QoL.	Minimize problems with item fit (IRT).
Avoid items for which more is not always better.	Minimize disordered thresholds (IRT).
Captures domain of interest.	Minimize DIF (IRT).
1F. Suitable for valuation	3D. Acceptable to stakeholders
Can be traded off.	Consider removal if disliked by patients in PPIE.
Amenable to change through health interventions.	3E. Reduce redundancy/overlap
1G. Suitable for translation	Competing items retain more general vs specific.

Abbreviations: CMQM = Comprehensive Model of QoL in Muscular Dystrophy; DIF = differential item functioning; DMD = Duchenne muscular dystrophy; EFA/CFA = exploratory/confirmatory factor analysis; IRT = Item response theory; MSA = Mokken scale analysis; PPIE = patient and public involvement and engagement; QoL = quality of life. Stage 1 rules adapted from the EQALY project.<sup>15</sup>

lead researcher, with 5 of the remaining transcripts secondary coded. Throughout coding, the researchers kept track of proposed iterative modifications to the framework based on the data. After all transcripts had been analyzed, the researchers met again to discuss and agree on a final framework for item generation.

The results of the analysis were discussed and refined with stakeholders during a PPIE advisory meeting (with patients and patients’ parents). Initial draft items for the DMD-QoL were then generated from the themes by the research team, in combination with advisory discussions with consultant NHS clinicians (including consultants in pediatric neurology, adult

neuromuscular disorders, and physiotherapy). Draft items were developed following a set of rules developed in previous projects while bearing in mind the scope of the current measure (table 2).<sup>15</sup> This included consideration of the minimum age of the target population (7 years) and wording items to be reading level appropriate.

### Stage 2: Initial Item Selection

To assess the content validity of the draft DMD-QoL and to reduce redundant items, draft items were shown to patients (n = 10), patients' parents (n = 10), and clinicians (n = 8) for their feedback. This sample size was based on a recommended sample size of ≈30 participants for pretesting questionnaires.<sup>16</sup> All participants were based in the United Kingdom. The parents were predominantly mothers (70%) 30 to 59 years of age. Likewise, the clinicians interviewed were largely female (75%) between 40 and 59 years of age. Background characteristics for the patients are given in table 3. Eligible parents and clinicians, who either were a parent of someone with DMD or worked with people with DMD, were opportunity sampled via Duchenne UK. The patients were all participants in the stage 1 qualitative study, so this provided an opportunity for "member checking" our analytical decisions.<sup>5,17</sup>

In a face-to-face or online interview, participants provided feedback on the 43 draft items, which were grouped (color-coded) by theorized qualitative theme. The participants received the set of items in advance. These cognitive debriefing interviews took place between February 2019 and April 2019. They were not recorded, but anonymized notes were taken. The exercise was designed to assess relevance (are the items, response options, and recall period appropriate?), comprehensiveness (are all key concepts included?), and comprehensibility (are the PROM instructions, items, and response options understood as intended?).<sup>18</sup> In the case of overlapping items (e.g., "I felt sad" vs "I felt happy"), participants were asked their preference. A topic guide (available online: doi.org/10.15131/shef.data.12942998.v1, file B) was produced to help facilitate this cognitive debriefing exercise.

### Stage 3: Final Item Selection and Psychometric Survey

To make a final item selection for the DMD-QoL, 3 sources of data were used, combined with a set of item selection rules (table 2). First, a stakeholder exercise had 3 groups (PPIE patient/parent representatives, clinicians, or health economist researchers) sorting the items within each qualitative domain into 1 of 3 outcomes. A traffic-light system was used (red = should not be in the final measure, amber = undecided or overlaps with another item, green = should be in the final measure), with the requirement that at least 1 item was in the green category per domain. Second, the items underwent an initial pretranslatability assessment by Oxford University Innovation Clinical Outcomes (including Afrikaans, French, German, Hebrew, Hindi, Russian, Simplified Chinese, and Spanish) to highlight any translatability concerns. Third, a

**Table 3** Patient Background Characteristics for Stage 1 Interviews and Stage 2 Cognitive Debriefing

	Stage 1 interview sample, n	Stage 2 debrief sample, n
<b>Age group, y</b>		
7-9	2	2
10-14	6	3
15-19	3	1
20-29	4	2
30-39	2	1
≥40	1	1
<b>Assistance needed to walk</b>		
Able to walk for distances >10 min without a mobility device	2	2
Able to walk for distances <10 min without a mobility device	5	2
Always uses a mobility device	11	6
<b>Assistance needed to transfer (from a bed to a chair)</b>		
Can transfer independently without assistance	6	4
Can transfer with some support	1	0
Needs a person or lifting device to transfer	11	6
<b>Assistance needed to eat</b>		
Can raise own hand to mouth with food or drink independently	9	6
Can raise own hand to mouth with food or drink with some support	2	0
Needs another person to bring food or drink to mouth	7	4
<b>Assistance needed to breathe</b>		
Does not use a ventilator at all	14	8
Uses a ventilator at night	1	0
Uses a ventilator during the day and night	3	2
<b>Use of steroid and heart medication</b>		
Does not take steroid or heart medication	1	1
Takes steroid medication	3	3
Takes heart medication	3	1
Takes steroid and heart medication	11	5
<b>Total, n</b>	<b>18</b>	<b>10</b>

psychometric survey was conducted with >100 UK DMD patients and caregivers.

## Psychometric Survey

The minimum sample size for the psychometric survey was 100 participants, given that numbers above this tend to provide more stable estimates in item response theory (IRT) analyses.<sup>19</sup> UK patients and parents were recruited between August 2019 and January 2020 via closed recruitment channels by the participating NHS sites and charities and support groups (including Duchenne UK, Muscular Dystrophy UK, and DMD Pathfinders). Eligibility criterion was a self-disclosed diagnosis of DMD (or to be responding on behalf of someone with a diagnosis of DMD). Participants could request a £5 Amazon e-voucher for taking part. To facilitate inclusivity, the survey could be completed by self-report (by a patient), assisted self-report (someone helped the patient complete the survey), or proxy (someone answered on behalf of the patient). It was made clear that only 1 response was allowed per patient. The survey was completed online and hosted on Qualtrics.

Participants completed several background and clinical questions on physical functioning, including the Brooke scale,<sup>20</sup> medication and treatment, and age. They then completed, in a fixed order, a 27-item draft DMD-QoL questionnaire (self-report or proxy), EQ-5D with age-appropriate variants (either EQ-5D-5L or EQ-5D-Y,<sup>8,21</sup> self-report or proxy), and the Pediatric Quality of Life Inventory (PedsQL) Generic Core Scales (GCS)<sup>22</sup> with age-appropriate variants: PedsQL GCS young child (ages 5–7 years), proxy only; and PedsQL GCS child (ages 8–12 years), adolescent (ages 13–17 years), young adult (ages 18–25 years), and adult (ages ≥26 years), all self-report or proxy. A copy of the survey outline is available online (doi.org/10.15131/shef.data.12942998.v1, file C). The survey took a median of 11.7 minutes to complete.

Psychometric analyses conducted on the items included Mokken scale analysis (MSA), IRT analyses using a rating scale model, and factor analyses. MSA was used to examine for violations of homogeneity of scale ( $H < 0.30$ ) and significant violations of local independence, monotonicity, and invariant item ordering.<sup>23</sup> IRT was used to explore violations of item fit using mean squares (expected range 0.6–1.4)<sup>23</sup> and ordered thresholds. Items were also assessed for differential item functioning (DIF) based on report method (self-report or assisted self-report vs proxy) and adult (>15 years) or child (7–15 years) status using a default, recommended  $\alpha$  level of 0.01.<sup>24</sup> Sixteen years of age was used as the cutoff point for adulthood in the current research because it is where children begin to transition to adult services in the United Kingdom. Analyses were conducted independently on unidimensional subscales (wherever possible) and, when not possible, on the total QoL scale.

Initial confirmatory factor analyses (CFAs) were conducted according to underlying theory. The qualitative themes were too numerous to fit to the data, and some of the themes had <3 indicators. Accordingly, to facilitate the CFA, items were

mapped to the Comprehensive Model of QoL in Muscular Dystrophy (CMQM).<sup>2</sup> This model has 3 domains: physical, psychological, and social. Details of this mapping are available online (doi.org/10.15131/shef.data.12942998.v1, file D). Given that the 3 QoL domains were assumed to be correlated and a total score was to be generated for the PROM, a hierarchical CFA was modeled, with items loading onto the 3 subscales and the 3 subscales loading onto an overall QoL factor. A unidimensional CFA was also modeled for comparison. Follow-up exploratory factor analyses (EFAs) were conducted on a baseline 1-factor and 2- and 3-factor solutions (as suggested as the potential optimal number of factors by factor diagnostics, including very simple structure,<sup>25</sup> the Velicer minimum average partial test,<sup>26</sup> and parallel analysis) on the polychoric correlation matrix using oblimin rotation due to poor CFA fit.

## Scoring and Relationship With Other Variables

After the final item selection, scores on the DMD-QoL were calculated and correlated against known sociodemographic and clinical characteristics, EQ-5D score, and PedsQL GCS score. The PedsQL GCS and its subscales were scored in a summative way, taking the mean across items within each scale. Utility scores for the EQ-5D-Y were generated with the 3L value set,<sup>27</sup> while the EQ-5D-5L was scored with the recommended cross-walk algorithm.<sup>28</sup> EQ-5D-Y and EQ-5D-5L responses were combined in the analyses because the utility scores are generated using the same value set.

All analyses were conducted in R x64 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria)<sup>29</sup> with the DescTools,<sup>30</sup> eRm,<sup>31</sup> lavaan,<sup>32</sup> lordif,<sup>24</sup> mokken,<sup>33</sup> and psych packages.<sup>34</sup>

## Data Availability

Anonymized data supporting the results of the psychometric survey are available online (doi.org/10.15131/shef.data.12942776.v1).

## Results

### Stage 1: Item Generation

Seven higher-order themes were extracted from the qualitative data, of which 6 were directly relevant to QoL and 1 (health care and support) was deemed as a process attribute (relating to treatment and care), rather than QoL. The QoL themes were physical aspects, social relationships, autonomy and independence, identity, feelings and emotions, and daily activities. Example quotes for these themes are available online (doi.org/10.15131/shef.data.12942998.v1, file I). A total of 43 draft items were generated for the DMD-QoL, which mapped onto these 6 themes. The full list of draft items and how they mapped onto the qualitative themes is available online (doi.org/10.15131/shef.data.12942998.v1, file E). These draft items intentionally included multiple questions on the same concept that differed linguistically for testing at cognitive debriefing (e.g., “I felt sad” vs “I felt unhappy”). A 4-item response scale and 1-week recall period were selected

on the basis of a review of existing measures, consideration of the PROM age range, and PPIE and clinician input. Both of these decisions were tested at cognitive debriefing. Language for the response options (chosen to be on a frequency scale) was generated with consideration of how often common response option terms were used by participants during the interviews (e.g., “sometimes” was used by all participants). The response options selected were “never,” “sometimes,” “a lot of the time,” and “all of the time.”

## Stage 2: Initial Item Selection

Minor modifications were made to the PROM instructions after a cognitive debriefing, including a “prefer not to disclose” option (for use in the psychometric survey only) and the use of “he” as opposed to “they” on the proxy version. While there were mixed responses on a 4-point response scale and a 1-week recall period, the consensus was to leave these as they were, particularly for a scale that was designed for young children. Participants had mixed responses on item ordering but preferred physical items first, before emotional items. Because of some difficulties in understanding the questions in participants <10 years old, we recommend that proxy report be used for children between 7 and 10 years of age.

Regarding comprehensiveness of the PROM, participants made a number of suggestions for additional content, of which 7 items were taken forward into the revised PROM, including questions on breathing and eating. Twenty-three items were dropped from the draft PROM (18 of these were redundant or overlapping items), resulting in a revised 27-item version. Summarized results from the cognitive debriefing and the rationale for keeping and dropping individual items at this stage are available in an item-tracking matrix online (available online: doi.org/10.15131/shef.data.12942998.v1, file F).

## Stage 3: Final Item Selection

Discussions within the patient and parent PPIE group revealed clear preferences for a subset of the 27 items, with 14 coded as green, 6 as amber, and 7 as red. The translatability assessment suggested 1 problematic item (“I was happy with the people around me”) and minor modifications to the instructions.

From a total of 225 people who accessed the psychometric survey, 107 answered the DMD-QoL, and 118 dropped out of the survey before answering the PROM. Of these 107, 5 people were excluded (2 people took >24 hours to complete the survey, 2 gave clinically implausible answers for someone with DMD according to published age-based norms and in consultation with a clinician,<sup>35</sup> and 1 responded to <80% of the DMD-QoL items). This resulted in a valid sample for analysis of 102. Overall 1.51% of data on the DMD-QoL were missing (i.e., answered as “prefer not to disclose”) and were median imputed to maximize data available for analysis.

Participant health and sociodemographic characteristics are given in table 4. Thirty-seven responses were self- or assisted self-reported, and 66 responses were proxy reported. Patients’ ages

ranged from 7 to 44 years (mean 15.77 years, SD 7.87 years). A full distribution of responses on the 27 item set is available online (doi.org/10.15131/shef.data.12942998.v1, file G).

The initial psychometrics on the full 27-item version revealed a poor fit to the CMQM model:  $\chi^2(321) = 735.700$ ,  $p < 0.001$ ; CFI = 0.699; root mean square error of approximation (RMSEA) = 0.113 (0.102, 0.123),  $p < 0.001$ , or a unidimensional model,  $\chi^2(324) = 893.40$ ,  $p < 0.001$ ; CFI = 0.549; RMSEA = 0.146 (0.134, 0.157),  $p < 0.001$ . Correspondingly, MSA revealed problems with homogeneity (4 items), local independence (2 items), and invariant item ordering (6 items). The IRT analysis revealed problems with MSQ item fit (4 items) and disordered thresholds (3 items). Full psychometric results are available online (doi.org/10.15131/shef.data.12942998.v1, file D).

To move toward a parsimonious measure (reducing overlap and patient burden), the following rules were applied: an item disliked by patients in the PPIE group was a candidate for removal; items needed to be amenable to change via health care intervention; more general items were preferred over specific items; and at least 1 item was to be retained per original qualitative domain and at least 3 items were to be retained per CMQM domain. This resulted in a reduction from 27 items to 15 items. Decisions on which items were dropped/retained and reasons why are in an item-tracking matrix available online (doi.org/10.15131/shef.data.12942998.v1, file H).

Analysis of the 15 items revealed a marginal improvement in fit for the 3-domain CMQM model,  $\chi^2(87) = 190.33$ ,  $p < 0.001$ ; CFI = 0.790; RMSEA = 0.108 (0.087, 0.129),  $p < 0.001$ , but still not within acceptable levels. Problems were also detected in the MSA and IRT. Accordingly, 1-factor, 2-factor, and 3-factor EFAs were conducted to help revise the model (doi.org/10.15131/shef.data.12942998.v1, file D). The EFA suggested an alternative latent structure for some of the items, which also made theoretical sense. As a result, the measurement model was revised, with 3 items (“I was in pain,” “I felt tired,” and “I found it hard to talk to people”) reallocated to the psychological theme, 2 items (“I found it hard to get around” and “I could take part in the things I wanted to”) moved to the social (participation) theme, and 1 item (“I felt left out”) dropped due to cross-loading (>0.3) on the EFA. This resulted in a 14-item PROM across 3 domains.

Psychometrics on the final 14-item PROM revealed a well-fitting (albeit not confirmatory) 3 domain hierarchical model,  $\chi^2(74) = 81.91$ ,  $p = 0.247$ ; CFI = 0.982; RMSEA = 0.032 (0.000, 0.067),  $p = 0.761$ , which fit better than a unidimensional model,  $\chi^2(77) = 197.75$ ,  $p < 0.001$ ; CFI = 0.727; RMSEA = 0.124 (0.103, 0.145),  $p < 0.001$ . The resultant conceptual model for the DMD-QoL is presented in figure 2. No problems were detected in MSA. Some minor problems remained in MSQ item fit (4 items) at the 0.6 to 1.4 threshold (but none so large as to degrade the measurement system<sup>36</sup>), and the physical functioning subscale had disordered

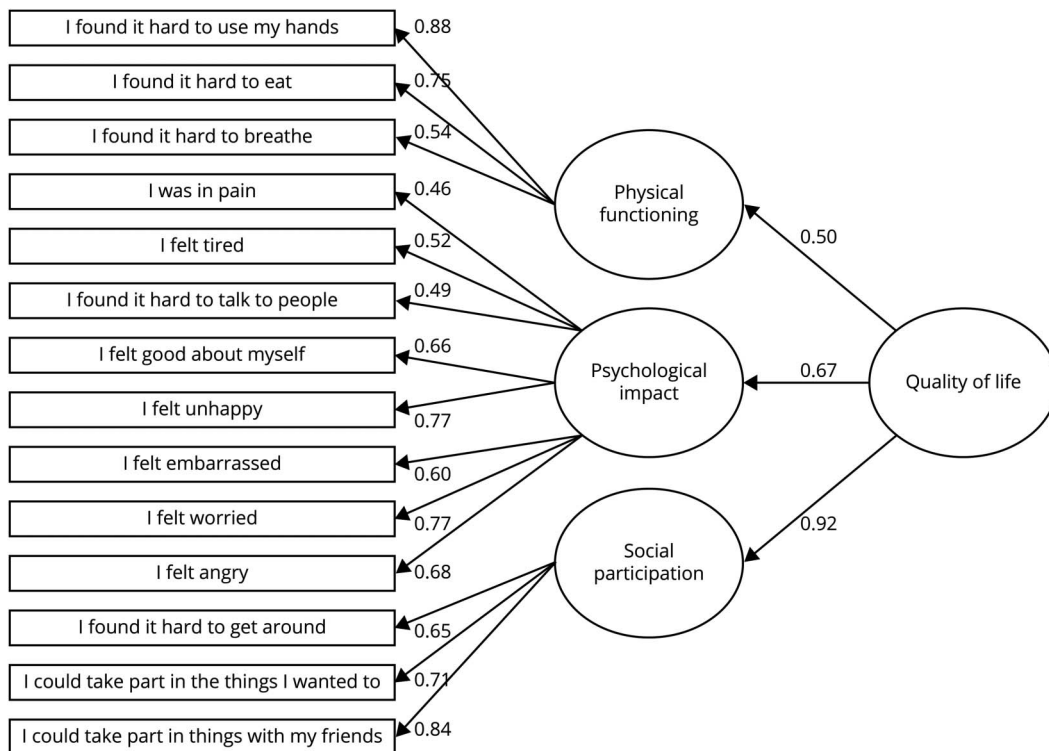
**Table 4** Health and Sociodemographic Characteristics for the Stage 3 Psychometric Survey Sample

Which of the following can you do independently (without any help)?	Yes, n			Yes (partially or with help), n	No, n	NA, n
Rise from lying on the floor to stand	2			27	69	4
Stand up from a chair	13			30	59	0
Stay standing for 3 s	37			12	3	0
Climb 4 stairs	7			28	67	0
Transfer from a bed to a chair	33			15	54	0
Walk/run for 10 m	Run			Run/walk	Walk	No
	13			14	16	59
Brooke score	1	2	3	4	5	6
	56	4	13	4	14	9
Medication use	Yes			No, but I used to	No, I never have	
Steroids	68			17	17	
Heart medication	62			3	37	
Treatments	Yes			No		
Treatment for weaker bones	33			69		
Spinal fusion surgery	16			86		
Feeding tube	6			96		
Ventilation	Yes, day and night			Yes, night		No
	11			12		79
Comorbid conditions	ASD	AD(H)D	OCD	Learning disability	None	NA
	21	6	4	23	61	2
QoL measure	Minimum		Maximum	Mean	SD	
DMD-QoL total	9		41	27.77	6.23	
DMD-QoL physical functioning	1		9	7.37	1.94	
DMD-QoL social participation	0		9	4.46	2.15	
DMD-QoL psychological impact	1		23	15.94	3.88	
PedsQL GCS total <sup>a</sup>	11.96		84.78	45.07	14.98	
PedsQL physical functioning <sup>a</sup>	3.13		90.63	26.41	22.12	
PedsQL emotional functioning <sup>a</sup>	10		100	56.61	20.51	
PedsQL social functioning <sup>a</sup>	10		90	48.52	19.57	
PedsQL school/work functioning <sup>a</sup>	5		100	60.42	22.00	
EQ-5D-Y <sup>b</sup>	-0.594		0.814	0.309	0.401	
EQ-5D-5L <sup>c</sup>	-0.268		0.850	0.200	0.270	
EQ-5D combined <sup>d</sup>	-0.594		0.850	0.206	0.346	

Abbreviations: AD(H)D = attention-deficit (hyperactivity) disorder; ASD = autism spectrum disorder; PedsQL GCS = Pediatric Quality of Life Inventory Generic Core Scales; NA = missing data (answered as "prefer not to disclose"); OCD = obsessive compulsive disorder; QoL = quality of life. Subscale scores not calculated for the EQ-5D as per standard practice. N = 102 (except as indicated: <sup>a</sup>n = 96, <sup>b</sup>n = 26, <sup>c</sup>n = 31, <sup>d</sup>n = 57). Full copies of the questions used in the survey are included in supplementary file C (doi.org/10.15131/shef.data.12942998.v1).



**Figure 2** Conceptual Model of the DMD-QoL



Standardized factor loadings shown (N = 102). DMD = Duchenne muscular dystrophy; QoL = quality of life.

thresholds (but not when incorporated as a total scale), which could potentially be remedied by reducing the response options to 3 for this subscale. No DIF was detected for reporting method. While 2 of the subscales could not be assessed for DIF (<4 items), 1 item in the psychological subscale (“I felt tired”) displayed adult/child DIF, with children typically reporting feeling more tired than adults. This is potentially explicable in that children tend to be still mobile, but their weak musculature results in extra effort, and they will experience fatigue as a result of this.<sup>37</sup> When considering the scale as a whole, 5 additional items displayed potential adult/child DIF, 4 of which (“I found it hard to get around,” “I found it hard to use my hands” “I found it hard to eat,” and “I found it hard to breathe”) are due to expected underlying differences in ability with age and thus represent benign rather than adverse DIF.<sup>38</sup> The other item (“I could take part in things with my friends”) displayed uniform DIF, with children more likely to endorse this item than adults. This suggests that it may be interpreted differently across age groups but is limited by analysis of a multidimensional item set with a relatively small sample size.<sup>39</sup> Accordingly, no further changes were considered justified at this stage in the absence of further validation in an independent dataset.

### Scoring and Relationship With Other Variables

Summative scores were calculated for the 3 DMD-QoL subdomains and the total scale score (supported by the hierarchical CFA), with 3 items reversed so that a higher score represented a better QoL (3 = never, 0 = all the time). The

Cronbach  $\alpha$  for the total DMD-QoL was 0.84, with  $\alpha$  values of 0.76, 0.74, and 0.83 for the physical, social, and psychological subscales, respectively. Summary scores for the QoL measures are listed in table 4.

For the DMD-QoL, the average total score was the same for adults (mean 27.6, SD 6.4) and children (mean 27.9, SD 6.2), although this differed across the subscale scores. Adults scored significantly lower on the physical subscale (mean 6.6, SD 2.2) than children (mean 8.0, SD 1.5),  $t(71.38) = 3.45, p < 0.001$ . Whereas adults scored significantly higher on the psychological subscale (mean 16.9, SD 3.5) than children (mean 15.2, SD 4.0),  $t(97.8) = 2.22, p = 0.029$ , on the social subscale, scores for adults (mean 4.1, SD 2.3) and children (mean 4.7, SD 2.0) did not significantly differ. Reported QoL was lowest for assisted report (mean 26.7, SD 7.7), followed by proxy report (mean 27.4, SD 5.5) and self-report (mean 30.0, SD 7.0), on the DMD-QoL. None of these differences were statistically significant.

EQ-5D utility score was significantly higher in children (mean 0.309, SD 0.401) than adults (mean 0.120, SD 0.270),  $t(42.55) = 2.04, p = 0.047$ , although this is generated with different measures. As with the DMD-QoL, lowest utilities were reported for the assisted report (mean -0.019, SD 0.091). However, utilities were lower in the self-report (mean 0.175, SD 0.286) than proxy report (mean 0.279, SD 0.392) group. The differences between groups were significant at the

**Table 5** Pearson or Point-Biserial Correlations Between QoL Measures and Background Characteristics in Stage 3 Psychometric Survey

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) DMD-QoL physical	—											
(2) DMD-QoL social	0.37 <sup>a</sup>	—										
(3) DMD-QoL psychological	0.25 <sup>a</sup>	0.51 <sup>a</sup>	—									
(4) DMD-QoL total	0.59 <sup>a</sup>	0.78 <sup>a</sup>	0.88 <sup>a</sup>	—								
(5) EQ-5D-Y utility score	0.46 <sup>a</sup>	0.52 <sup>a</sup>	0.65 <sup>a</sup>	0.68 <sup>a</sup>	—							
(6) EQ-5D-5L utility score	0.60 <sup>a</sup>	0.49 <sup>a</sup>	0.22	0.48 <sup>a</sup>	NA	—						
(7) EQ-5D combined utility score	0.54 <sup>a</sup>	0.50 <sup>a</sup>	0.40 <sup>a</sup>	0.57 <sup>a</sup>	1.00 <sup>a</sup>	1.00 <sup>a</sup>	—					
(8) PedsQL physical	0.34 <sup>a</sup>	0.30 <sup>a</sup>	0.20 <sup>a</sup>	0.33 <sup>a</sup>	0.46 <sup>a</sup>	0.21	0.41 <sup>a</sup>	—				
(9) PedsQL emotional	0.15	0.45 <sup>a</sup>	0.76 <sup>a</sup>	0.67 <sup>a</sup>	0.60 <sup>a</sup>	0.15	0.30 <sup>a</sup>	0.25 <sup>a</sup>	—			
(10) PedsQL social	0.04	0.50 <sup>a</sup>	0.65 <sup>a</sup>	0.58 <sup>a</sup>	0.49 <sup>a</sup>	0.24	0.25 <sup>b</sup>	0.24 <sup>a</sup>	0.50 <sup>a</sup>	—		
(11) PedsQL school/work	-0.05	0.36 <sup>a</sup>	0.46 <sup>a</sup>	0.39 <sup>a</sup>	0.54 <sup>a</sup>	0.12	0.21	0.11	0.52 <sup>a</sup>	0.55 <sup>a</sup>	—	
(12) PedsQL total	0.22 <sup>a</sup>	0.55 <sup>a</sup>	0.66 <sup>a</sup>	0.66 <sup>a</sup>	0.66 <sup>a</sup>	0.31	0.48 <sup>a</sup>	0.69 <sup>a</sup>	0.74 <sup>a</sup>	0.73 <sup>a</sup>	0.69 <sup>a</sup>	—
(13) Age	-0.52 <sup>a</sup>	-0.20 <sup>a</sup>	0.09	-0.17 <sup>b</sup>	-0.47 <sup>a</sup>	-0.39 <sup>a</sup>	-0.43 <sup>a</sup>	-0.36 <sup>a</sup>	0.03	0.20 <sup>b</sup>	0.37 <sup>a</sup>	0.00
(14) Can rise from floor	0.33 <sup>a</sup>	0.27 <sup>a</sup>	0.07	0.24 <sup>a</sup>	0.56 <sup>a</sup>	0.46 <sup>a</sup>	0.57 <sup>a</sup>	0.44 <sup>a</sup>	0.18 <sup>b</sup>	0.10	-0.08	0.29 <sup>a</sup>
(15) Can stand from chair	0.44 <sup>a</sup>	0.32 <sup>a</sup>	0.04	0.27 <sup>a</sup>	0.72 <sup>a</sup>	0.61 <sup>a</sup>	0.67 <sup>a</sup>	0.43 <sup>a</sup>	0.04	0.03	-0.11	0.21 <sup>a</sup>
(16) Can stay standing	0.48 <sup>a</sup>	0.25 <sup>a</sup>	-0.02	0.22 <sup>a</sup>	0.58 <sup>a</sup>	0.66 <sup>a</sup>	0.64 <sup>a</sup>	0.27 <sup>a</sup>	-0.01	-0.04	-0.15	0.09
(17) Can walk/run 10 m	0.41 <sup>a</sup>	0.28 <sup>a</sup>	0.07	0.27 <sup>a</sup>	0.74 <sup>a</sup>	0.76 <sup>a</sup>	0.70 <sup>a</sup>	0.43 <sup>a</sup>	-0.01	0.03	-0.10	0.20 <sup>b</sup>
(18) Can climb 4 stairs	0.40 <sup>a</sup>	0.29 <sup>a</sup>	0.07	0.27 <sup>a</sup>	0.70 <sup>a</sup>	0.61 <sup>a</sup>	0.69 <sup>a</sup>	0.48 <sup>a</sup>	0.07	0.07	-0.07	0.27 <sup>a</sup>
(19) Can transfer bed to chair	0.49 <sup>a</sup>	0.25 <sup>a</sup>	0.03	0.26 <sup>a</sup>	0.72 <sup>a</sup>	0.62 <sup>a</sup>	0.69 <sup>a</sup>	0.32 <sup>a</sup>	0.02	-0.01	-0.11	0.14
(20) Brooke score	-0.75 <sup>a</sup>	-0.26 <sup>a</sup>	0.04	-0.30 <sup>a</sup>	-0.51 <sup>a</sup>	-0.60 <sup>a</sup>	-0.56 <sup>a</sup>	-0.24 <sup>a</sup>	0.07	0.18 <sup>b</sup>	0.23 <sup>a</sup>	0.01
(21) Takes steroids	0.48 <sup>a</sup>	0.12	-0.10	0.13	-0.01	0.47 <sup>a</sup>	0.28 <sup>a</sup>	0.10	-0.06	-0.20 <sup>a</sup>	-0.20 <sup>b</sup>	-0.08
(22) Takes heart medication	-0.19 <sup>b</sup>	-0.07	0.05	-0.05	-0.13	-0.17	-0.24 <sup>b</sup>	-0.38 <sup>a</sup>	0.05	0.02	0.11	-0.15
(23) Had bone treatment	0.09	0.15	-0.06	0.05	-0.44 <sup>a</sup>	0.39 <sup>a</sup>	-0.07	-0.04	0.08	-0.04	0.03	0.01
(24) Had spinal fusion	-0.53 <sup>a</sup>	-0.13	0.07	-0.17 <sup>b</sup>	NA	-0.49 <sup>a</sup>	-0.38 <sup>a</sup>	-0.07	0.02	0.14	0.22 <sup>a</sup>	0.06
(25) Uses electric wheelchair	-0.47 <sup>a</sup>	-0.22 <sup>a</sup>	-0.02	-0.23 <sup>a</sup>	-0.58 <sup>a</sup>	-0.44 <sup>a</sup>	-0.58 <sup>a</sup>	-0.32 <sup>a</sup>	0.03	-0.03	0.16	-0.12
(26) Uses ventilation	-0.51 <sup>a</sup>	-0.15	0.06	-0.18 <sup>b</sup>	NA	-0.49 <sup>a</sup>	-0.38 <sup>a</sup>	-0.26 <sup>a</sup>	0.09	0.02	0.25 <sup>a</sup>	-0.03
(27) Has a feeding tube	-0.39 <sup>a</sup>	-0.09	-0.02	-0.17	NA	-0.29	-0.24 <sup>b</sup>	-0.19 <sup>b</sup>	-0.04	0.05	0.09	-0.06
(28) Has a comorbid condition	0.13	-0.06	-0.29 <sup>a</sup>	-0.16	-0.39 <sup>b</sup>	0.21	-0.09	0.03	-0.25 <sup>a</sup>	-0.33 <sup>a</sup>	-0.41 <sup>a</sup>	-0.30 <sup>a</sup>

Abbreviations: DMD = Duchenne muscular dystrophy; NA = no variation in at least one of the variables; PedsQL = Pediatric Quality of Life Inventory; QoL = quality of life.

N = 102. Correlations use pairwise complete observations. Full copies of the questions used in the survey are available online ([doi.org/10.15131/shef.data.12942998.v1](https://doi.org/10.15131/shef.data.12942998.v1), file C).

<sup>a</sup>  $p < .05$ .

<sup>b</sup>  $p < .10$ .

10% level,  $F_{2,54} = 2.611$ ,  $p = 0.083$ . Finally, for the PedsQL GCS total score, there was a nonsignificant trend for QoL to be lower in children (mean 44.1, SD 16.7) than adults (mean 46.4, SD 12.3). Regarding reporting method, self-report had the highest score (mean 50.3, SD 9.5), followed by assisted (mean 43.9, SD 14.2), and proxy (mean 43.7, SD 16.3) report. These differences were nonsignificant.

Pearson and point-biserial correlations between the QoL PROMs and background measures are displayed in table 5. The DMD-QoL total score had large significant positive correlations with the PedsQL GCS total score and the EQ-5D utility score, indicating convergent validity. Further convergent validity for the DMD-QoL physical functioning subscale was shown by its large significant negative correlation with age

and association with all other physical functioning background questions such as a large significant negative relationship with the Brooke score. The DMD-QoL social participation subscale had smaller significant correlations with the physical functioning questions, as would be expected, but not questions on medication, demonstrating some discriminant validity. Finally, the DMD-QoL psychological subscale had a small significant negative correlation with having a comorbid condition such as autistic spectrum disorder.

## Discussion

Our primary objective was to develop a new QoL PROM for use in boys and men with DMD (age  $\geq 7$  years) and to assess its preliminary psychometric properties. The DMD-QoL is a 14-item measure with self-report (age  $\geq 10$  years) and proxy (age  $\geq 7$  years) forms. The measure has a hierarchical 3-domain structure, with physical, social, and psychological subdomains, and a superordinate QoL factor. The DMD-QoL uses a summative scoring system for these domains. The DMD-QoL has evidenced content validity (from stages 1 and 2 of this developmental work) and preliminary good psychometric properties (with no problems in MSA and only minor problems observed with item fit, disordered thresholds, and DIF). The DMD-QoL is available for use from the licensors, Oxford University Innovation Clinical Outcomes.

The DMD-QoL development represents several successes, including the realization of rigorous PROM development in a rare condition (dealing with small numbers of patients), a high degree of PPIE and clinician input, and the inclusion of iterative qualitative work, which has resulted in a PROM with evident content validity. The DMD-QoL is a PROM designed for use across the life course, which has advantages in being able to compare differences in QoL across DMD progression on a common scale. The PROM is intentionally short (14 item) to minimize patient burden (which is potentially high in DMD)<sup>2,40</sup> and to facilitate subsequent transition to a preference-based measure.<sup>41</sup> The DMD-QoL and its preference-based variant will have benefit in informing resource allocation decisions for health technologies (based on their effect on QoL) and may have potential benefit in clinic as a tool to facilitate conversations and potentially to inform care and treatment decisions.

Despite its strengths, there are several apparent limitations of the PROM in its current form, which may be addressed in future work. First, the DMD-QoL was designed to be sufficiently comprehensive to assess aspects of QoL important to people with DMD while not being too burdensome. This means that the PROM may not be exhaustive and that certain aspects of QoL such as sexual relationships and end-of-life issues, which may be important to adults with DMD,<sup>42,43</sup> were considered inappropriate to be included in a questionnaire to be used from childhood. Nevertheless, future work could consider “bolt-on” questions to the DMD-QoL for adults.

Second, while we provide evidence here for the content validity of the scale and some initial evidence for its construct validity, further work is required to assess the PROM for its psychometric performance in an independent sample, including, for example, test-retest reliability and responsiveness.<sup>44</sup> Because the underlying latent structure of the measure was adjusted with the use of the current dataset, the model requires confirmation (validation) on an independent dataset. Relatedly, given the relatively small sample size, self-report and proxy report responses were combined in the psychometric analyses of the items and to inform item selection. While we believe this is defensible because the 2 variants were hypothesized to share the same measurement model and no DIF was detected as a function of reporting method, future confirmatory work may consider testing the self-report and proxy variants independently in larger, targeted samples. This is particularly the case for elements of QoL that are more difficult to observe such as in the psychological subscale.<sup>6</sup>

Third, the PROM was developed for use in the United Kingdom. The initial translatability assessment of the PROM looks promising, but it is not yet known whether the PROM will perform well in other countries and cultures. Fourth, the development of the DMD-QoL potentially suffers from selection biases in recruitment. Frequently, opportunity samples were used, and there is a need to further validate the measure in an independent sample with broader characteristics (e.g., in those unaffiliated with Duchenne UK; recruitment and participation in clinic rather than on an online survey). It is not yet known whether the validity of the PROM generalizes to the full population of people with DMD in the United Kingdom or indeed internationally, and further research is needed. This includes research in specific clinical groups such as those with comorbid conditions resulting in relatively lower cognitive function. Furthermore, limited background data were collected on the parents of patients with DMD taking part in the research. For example, we did not collect data on whether participating mothers were manifesting carriers. Characteristics such as these may have influenced responses, but this is not testable in the current data.

The 14-item DMD-QoL was designed in collaboration with, and for use in, the Duchenne community to have good content validity relative to existing QoL PROMs.<sup>5</sup> Initial analyses suggest that DMD-QoL performs well in people with DMD (and their caregivers). However, taking into account the aforementioned limitations, the DMD-QoL requires further independent validation and needs to be compared for its performance against alternative QoL instruments. In the next stage of its development, the measure will be valued to generate a preference-based measure that can be used to generate quality-adjusted life-years to inform resource allocation decisions within health technology assessments.

## Acknowledgment

The authors express their appreciation to all of the people with DMD and their parents who took the time to help with

this research. Their thanks go to Duchenne UK and the wider Project HERCULES consortium for their enduring support and feedback, including Josie Godfrey and Emily Crossley. They are grateful to all NHS collaborating clinicians and their teams who helped facilitate study recruitment and provided feedback on the work, including Anirban Majumdar, Anne Marie-Childs, Anna Mayhew, Stefan Spinty, Volker Staub, and Ros Quinlivan. Finally, their special thanks go to all of the PPIE representatives who collaborated on the project.

## Study Funding

This project was funded by the project HERCULES consortium. This project 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.

## Disclosure

P.A. Powell, J. Carlton, D. Rowen, F. Chandler, and J.E. Brazier report no disclosures relevant to the manuscript. M. Guglieri is an investigator in several commercial clinical trials (Sarepta, Italfarmaco, Pfizer, Santhera), study chair for ReveraGen (no funding received), received grants from Sarepta and 1 personal consultant payment from Sarepta for a talk, and is on the advisory board for Pfizer and NS Pharma. Go to [Neurology.org/N](http://Neurology.org/N) for full disclosures.

## Publication History

Received by *Neurology* November 25, 2020. Accepted in final form February 12, 2021.

## Appendix Authors

Name	Location	Contribution
<b>Philip A. Powell, PhD</b>	University of Sheffield, UK	Design and conceptualized study; acquired and analyzed the data; drafted the manuscript for intellectual content
<b>Jill Carlton, PhD</b>	University of Sheffield, UK	Design and conceptualized study; interpreted the data; revised the manuscript for intellectual content
<b>Donna Rowen, PhD</b>	University of Sheffield, UK	Design and conceptualized study; interpreted the data; revised the manuscript for intellectual content
<b>Fleur Chandler, MSc</b>	Duchenne UK, London	Interpreted the data; revised the manuscript for intellectual content
<b>Michela Guglieri, MD</b>	University of Newcastle and Newcastle Hospitals NHS Foundation Trust, UK	Interpreted the data; revised the manuscript for intellectual content
<b>John E. Brazier, PhD</b>	University of Sheffield, UK	Design and conceptualized study; interpreted the data; revised the manuscript for intellectual content

## References

1. Utley L, Carlton J, Woods HB, Brazier J. A review of quality of life themes in Duchenne muscular dystrophy for patients and carers. *Health Qual Life Outcomes* 2018;16:237.
2. Bann CM, Abresch RT, Biesecker B, et al. Measuring quality of life in muscular dystrophy. *Neurology* 2015;84:1034–1042.
3. Straub V, Mercuri E, DMD Outcome Measure Study Group. Report on the workshop: meaningful outcome measures for Duchenne muscular dystrophy, London, UK, 30–31 January 2017. *Neuromuscul Disord* 2018 28;690–701.
4. 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;18:263.
5. Powell PA, Carlton J, Rowen D, Brazier JE. Producing a preference-based quality of life measure for people with Duchenne muscular dystrophy: a mixed-methods study protocol. *BMJ Open* 2019;9:e023685.
6. US Department of Health and Human Services, Food and Drug Administration. Guidance for industry: patient reported outcome measures: use in medical product development to support labeling claims. 2009. Available at: [fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims](http://fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims). Accessed September 8, 2020.
7. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33:337–343.
8. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res* 2010;19:875–886.
9. Torrance GW, Feeny DH, Furlong WJ, 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.
10. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. *Pharmacoeconomics* 1995;7:490–502.
11. Stevens K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Qual Life Res* 2009;18:1105–1113.
12. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol* 2013;13:117.
13. Carlton J. Identifying potential themes for the child amblyopia treatment questionnaire. *Optom Vis Sci* 2013;90:867–873.
14. Oluboyede Y, Hulme C, Hill A. Development and refinement of the WAITE: a new obesity-specific quality of life measure for adolescents. *Qual Life Res* 2017;26:2025–2039.
15. Peasgood T, Mukuria C, Carlton J, Connell J, Brazier J. Criteria for item selection for a preference-based measure for use in economic evaluation. *Qual Life Res*. Epub 2020 Dec 7.
16. Perneger TV, Courvoisier DS, Hudelson PM, Gayet-Ageron A. Sample size for pretests of questionnaires. *Qual Life Res* 2015;24:147–151.
17. Lincoln YS, Guba EG. *Naturalistic Inquiry*. Sage; 1985.
18. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–1170.
19. Chen WH, Lenderking W, Jin Y, Wyrwich KW, Gelhorn H, Revicki DA. Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Qual Life Res* 2014;23:485–493.
20. Brooke MH, Griggs RC, Mendell JR, Fenichel GM, Shumate JB, Pellegrino RJ. Clinical trial in Duchenne dystrophy, I: the design of the protocol. *Muscle Nerve* 1981;4:186–197.
21. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–1736.
22. Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the Pediatric Quality of Life Inventory. *Med Care* 1999;37:126–139.
23. Dima AL. Scale validation in applied health research: tutorial for a 6-step R-based psychometrics protocol. *Health Psychol Behav Med* 2018;6:136–161.
24. 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;39:1.
25. Revelle WR, Thomas. Very simple structure: an alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behav Res* 1979;14:403–414.
26. Velicer WF. Determining the number of components from the matrix of partial correlations. *Psychometrika* 1976;41:321–327.
27. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;1095–1108.
28. 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;15:708–715.
29. A Language and Environment for Statistical Computing [computer program]. R Foundation for Statistical Computing; 2019.
30. Signorell A, Aho K, Alfons A, et al. DescTools: Tools for Descriptive Statistics: R Package Version 0.99.28. R Foundation for Statistical Computing; 2019.
31. Mair P, Hatzinger R. Extended Rasch modeling: the eRm package for the application of IRT models in R. *J Stat Softw* 2007;20:1–20.
32. Rosseel Y. Lavaan: an R package for structural equation modeling. *J Stat Softw* 2012;48:1–36.
33. Van der Ark LA. Mokken scale analysis in R. *J Stat Softw* 2007;20:1–19.
34. Revelle W. *Psych: Procedures for Personality and Psychological Research*. Northwestern University; 2018.

35. McDonald CM, Henricson EK, Abresch RT, et al. Long-term effects of glucocorticoids on function, quality of life, and survival in patients with Duchenne muscular dystrophy: a prospective cohort study. *Lancet* 2018;391:451–461.
36. Linacre J. What do infit and outfit, mean-square and standardized mean?. *Rasch Meas Trans* 2002;16:878.
37. Wei Y, Speechley KN, Zou G, Campbell C. Factors associated with health-related quality of life in children with Duchenne muscular dystrophy. *J Child Neurol* 2016;31:879–886.
38. Douglas JA, Roussos LA, Stout W. Item-bundle DIF hypothesis testing: identifying suspect bundles and assessing their differential functioning. *J Educ Meas* 1996;33:465–484.
39. Scott NW, Fayers PM, Aaronson NK, et al. Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health Qual Life Outcomes* 2010;8:81.
40. Squitieri L, Bozic KJ, Pusic AL. The role of patient-reported outcome measures in value-based payment reform. *Value Health* 2017;20:834–836.
41. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press; 2017.
42. Abbott D, Jepson M, Hastie J. Men living with long-term conditions: exploring gender and improving social care. *Health Soc Care Community* 2016;24:420–427.
43. Abbott D, Prescott H, Forbes K, Fraser J, Majumdar A. Men with Duchenne muscular dystrophy and end of life planning. *Neuromuscul Disord* 2017;27:38–44.
44. Mokkink LB, 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 outcomes. *J Clin Epidemiol* 2010;63:737–745.