UC Davis
UC Davis Previously Published Works

Title

Rasch Analysis of the Pediatric Quality of Life Inventory 4.0 Generic Core Scales Administered to Patients With Duchenne Muscular Dystrophy.

Permalink

https://escholarship.org/uc/item/8x11g4s7

Journal

Value in Health, 24(10)

Authors

Landfeldt, Erik
Iff, Joel
Henricson, Erik

Publication Date

2021-10-01

DOI

10.1016/j.jval.2021.05.016

Peer reviewed

Rasch Analysis of the Pediatric Quality of Life Inventory 4.0 Generic Core Scales Administered to Patients with Duchenne Muscular Dystrophy

Erik Landfeldt, PhD[1,2], **Joel Iff, PhD**[3], **Erik Henricson, PhD, MPH**[4], **CINRG Investigators**[*]

[1]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

[2]ICON plc, Stockholm, Sweden

[3]Sarepta Therapeutics, Inc, Cambridge, Massachusetts, USA

[4]Department of Physical Medicine & Rehabilitation, University of California Davis School of Medicine, Sacramento, CA, USA

Abstract

Objectives: The objective of this study was to examine the psychometric properties of the Pediatric Quality of Life Inventory 4.0 Generic Core Scales (PedsQL 4.0 GCS) in Duchenne muscular dystrophy (DMD), a rare, severely debilitating, and ultimately fatal neuromuscular disease.

Methods: Patients with DMD were recruited from 20 centres across nine countries as part of the Cooperative International Neuromuscular Research Group Duchenne Natural History Study (NCT00468832). The psychometric properties of the PedsQL 4.0 GCS were examined using Rasch analysis.

Results: In total, 329 patients with DMD (mean age: 9 years, range: 3–18 years; 75% ambulatory) completed the PedsQL 4.0 GCS. The most difficult instrument items, expressing the greatest loss in HRQoL, were those associated with emotional well-being (e.g., being teased by other children, feeling sad, and not making friends), as opposed to somatic disability (e.g., lifting heavy objects, participating in sports, and running). The mean item and person fit residuals were estimated at 0.301 (SD: 1.385) and –0.255 (1.504), respectively. In total, 87% (20 of 23) of items displayed disordered thresholds, and many exhibited non-trivial dependency. The overall item-trait interaction χ^2 value was 178 (115 degrees of freedom, p<0.001). Our analysis also revealed significant issues with differential item functioning, and by investigating residual principal component loadings, the PedsQL 4.0 GCS total score was found to be multidimensional.

Conclusions: The PedsQL 4.0 GCS records information clinically relevant to patients with DMD, but the total scale score may not be fit for purpose as a measure HRQoL in this disease population.

Précis:

The PedsQL 4.0 GCS, a common measure of health-related quality of life (HRQoL), may not be fit for purpose for use in Duchenne muscular dystrophy.

Keywords

Psychometric analysis; Patient reported outcome; Quality of Life; CINRG; Disability

1 Introduction

Duchenne muscular dystrophy (DMD) is a rare, X-linked, severely debilitating, and ultimately fatal neuromuscular disease characterised by progressive muscle weakness.[1] In recent decades, advances in the understanding of the cellular and molecular mechanisms delineating DMD have led to the discovery of several novel treatment strategies. Examples include stop codon readthrough, exon skipping, genome editing, utrophin modulation, and gene- addition therapies.[2] This acceleration in therapy development has resulted in a pressing need to identify patient-reported outcome scales that are fit for purpose to quantify drug benefits in randomized controlled trials (RCTs) to inform regulatory approval, as well as local decisions concerning pricing and reimbursement.[3]

One of the most commonly applied measures of health-related quality of life (HRQoL) in paediatric populations, including DMD, is the Pediatric Quality of Life Inventory 4.0 Generic Core Scales (PedsQL 4.0 GCS). The PedsQL 4.0 GCS is a multi-dimensional tool developed through focus groups and cognitive interviews with the aim to be applicable to healthy school and community populations, as well as paediatric populations with acute and chronic health conditions.[4] Yet, despite widespread use in both RCTs and observational research, evidence of the psychometric properties of the PedsQL 4.0 GCS in DMD is limited. To help bridge this evidence gap, the objective of our study was to explore the psychometric properties of the PedsQL 4.0 GCS administered to patients with DMD using Rasch analysis, a modern psychometric method complementary to traditional approaches based on classical test theory.[3]

2 Methods

2.1 The Pediatric Quality of Life Inventory 4.0 Generic Core Scales

The PedsQL 4.0 GCS was designed to measure the World Health Organization's core health domains (i.e., affect, cognition, pain, mobility, self-care, and usual activities) and an additional domain of school functioning.[4] The tool encompasses 23 items formulated as statements, each described in five levels (ranging from "Never" to "Almost Always"), covering four domains: (i) Physical Functioning (8 items), (ii) Emotional Functioning (5 items), (iii) Social Functioning (5 items), and (iv) School Functioning (5 items). The instrument is available in both self- and proxy-report versions in specific formats for ages

5–7, 8–12, and 13–18 years (formats for proxy-reports also contain a version for ages 2–5 years). The items within the PedsQL 4.0 GCS are scored using the Likert method of summated ratings and transformed to a scale ranging from 0 to 100, where a higher score indicates higher HRQoL.[4]

To date, only two studies have tested the psychometric properties of the PedsQL 4.0 GCS in samples of patients with DMD. Davis et al. found that the scale exhibited satisfactory psychometric properties with respect to feasibility, reliability, and validity, and concluded that the tool is "a reliable measure of disease-specific HRQoL in the DMD population and may be used as an outcome measure in clinical trials."[5] Subsequently, Lim et al. conducted a Rasch analysis of PedsQL 4.0 GCS domain scores in a sample of 63 children with DMD and found that items and participants fitted the Rasch model relatively well.[6] In contrast, Rasch analyses of the PedsQL 4.0 GCS administered to other populations (i.e., children with cancer, preschool children with refractive errors, and healthy children and adolescents) have indicated several issues with both total and domain scores.[7–9] Suboptimal scale performance has also been demonstrated for the PedsQL 3.0 Neuromuscular Module, a less frequently used module designed to complement the PedsQL 4.0 GCS in neuromuscular populations, in patients with DMD.[10]

2.2   Patients and Procedures

The PedsQL 4.0 GCS data analysed as part of this study was collected in the Cooperative International Neuromuscular Research Group (CINRG) Duchenne Natural History Study (DNHS) (NCT00468832). In brief, DNHS was a prospective, observational study of patients between 2 and 30 years of age with clinically confirmed DMD recruited between 2006 and 2016 from a total of 20 centres across nine countries (Argentina, Australia, Canada, India, Israel, Italy, Puerto Rico, Sweden, the United States) (detailed inclusion and exclusion criteria have been previously described.[11–13] In the DNHS, patients were asked to self-complete the PedsQL 4.0 GCS (among other measures), with or without help from a caregiver (e.g., a parent). Additionally, a set of demographic and clinical patient characteristics (listed in Table 1) was also recorded. Study ethical approval was obtained from Institutional Review Boards at each centre. All participants provided written informed consent (via parents/other legal guardians, as necessary).

2.3   Rasch Analysis

Rasch analysis is the formal testing of a scale against a mathematical model developed by Danish mathematician Georg Rasch.[14] In contrast to classical test theory, which forms the basis of traditional, ordinal scales, the Rasch measurements model satisfies the strict criteria of fundamental measurement (i.e., linearity, sample-free calibration, test-free measurement, and unidimensionality[15,16]) and therefore allows for invariant comparisons between respondents, that is, meaningful interpretation and comparison of mean instrument scores within and across populations. For this reason, Rasch analysis has become "the measurement standard for patient reported outcomes in general".[17] The Rasch measurement model was initially developed for binary response options and subsequently generalised to polytomous contexts.[18]

In brief, there are three main components to the theory of Rasch measurement. First, the response from person n to item i is governed by two factors only:

1. Person ability, θ_n (e.g. level of HRQoL); and

2. Item difficulty, δ_i (e.g. the level of HRQoL expressed by the item).

The probability that a person will affirm an item is a function of the distance between person ability and item difficulty, that is, $\theta_n - \delta_i$. In other words, in the case of HRQoL, this means that the probability that a patient will affirm an item is dependent on his or her level of HRQoL and the level of HRQoL expressed by the item. Alternatively, the probability can be expressed using natural logarithms:

$$0 \le e^{(\theta_n - \delta_i)} \le \infty$$

Or as an odds ratio:

$$0 \le \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}} \le 1$$

Accordingly, the probability that person n will affirm item i given ability θ_n and item difficulty δ_i is:

$$P_i(\text{ affirm };\theta_n) = \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}}$$

The second component of the Rasch model is a probabilistic form of the Guttman response pattern, which states that if a person affirms a task then there is a high probability that easier tasks will also be affirmed. In the case of a measure of HRQoL, this implies that a patient who states that he or she is able to perform a task indicative of relatively low functional impairment and high HRQoL (e.g., running) also would be able to perform a task indicative of relatively high impairment and low HRQoL (e.g. walking).

The third and last component to the theory of Rasch measurement is Rasch's criterion of invariance, where item locations can be estimated independently of the distribution of person locations on the continuum, which ensures that results for scales are sample independent and results for samples are scale independent.[15] Put differently, this means that the instrument is stable (i.e., not sample dependent) and the property being measured is stable at one point in time (i.e., not instrument dependent). Only Rasch measurement can test stability of instruments and people; other parameters in item response theory models render these estimates sample dependent.[19]

The Rasch analysis input comprises the patient-level instrument data. The Rasch analysis output consists of an interval-level scale or metric (logit scale) to which both respondents and items are located. In addition, Rasch analysis provides a unified approach to test several important measurement issues, including disordered thresholds which occurs when

respondents have difficulty discriminating between levels of an item given their ability (that is, when a specific level never is the most probable response to a question), and differential item functioning (which occurs when, at the same level of ability, response to a particular item differs by a factor, e.g., sex).[20,21]

2.4 Statistical Analysis

We fitted a Rasch partial credit model,[22] determined based on a likelihood-ratio test, to the PedsQL 4.0 GCS data using RUMM2030 (RUMM Laboratory, Perth, Australia). Individual item misfit was defined as a fit residual >|2.5| or a χ^2 Bonferroni-adjusted p-value <0.002174 (0.05/23).[20,21] We also analysed person fit to the Rasch model (defined as a fit residual >|2.5|), ordering of item response category thresholds (i.e., that respondents are able to differentiate between response categories), local item dependency (i.e., if a reply to one item predicts the reply to another item, defined as 0.20 above the mean residual correlation for all items), targeting (i.e., the match of the different ability levels estimated through the Rasch model with the ability levels observed in our sample), reliability (Person Separation Index [PSI],[21,23] indicating the possibility of the scale to differentiate between respondents at different levels of caregiver burden), differential item functioning (i.e., item stability) investigated through analysis of variance by disease stage (as defined in Table 1) (differences by early/late non-ambulatory were not explored due to the limited sample size for these strata) and steroid use (any lifetime exposure vs. no exposure), Bonferroni-adjusted p-value <0.000725 [0.05/69]), and unidimensionality through principal components analysis of the residuals (described in more details below). In the analysis of differential item functioning, each person was first assigned to a person factor group (e.g., ambulatory status) and classified by ability measure of the latent trait into a class interval (as part of fitting the Rasch model in RUMM2030). Then, for each item, the observation residuals were analysed with a two-way analysis of variance by factor and factor-class interval interaction. The presence of differential item functioning is indicated by statistically significant inter-person factor-group variance.[24]

Unidimensionality was tested using the method proposed by Smith,[25] which in brief involves examining the correlation between items and the first residual factor. These patterns are subsequently used to define two subsets of items (i.e., the positively and negatively correlated items), which are then employed to make separate person estimates. By using an independent t-test for the difference in these estimates for each person, the percentage of such tests outside the range –1.96 to 1.96 should not exceed 5%. Finally, a confidence interval for a binomial test of proportions is calculated for the observed number of significant tests, and this value should overlap the 5% expected value for the scale to be unidimensional.[21,26]

Sensitivity Analysis—The psychometric properties of the English-language version of the PedsQL 4.0 GCS, including participants from Australia and Canada, were explored in sensitivity analysis. We were not able to include the US sample in this subset, since an unknown proportion of these patients completed the Spanish-language version of the PedsQL 4.0 GCS.

3 Results

In total, 329 patients with DMD that participated in the DNHS completed the PedsQL 4.0 GCS in accordance with instructions. A summary of patient demographic and clinical characteristics is presented in Table 1. The distribution of replies to the PedsQL 4.0 GCS items is presented in Figure 1.

3.1 Item Fit to the Rasch Model

Table 2 presents the fit of the PedsQL 4.0 GCS items to the Rasch model, ordered by item difficulty. Only item 14 ("Getting along with other children") displayed model misfit (classic under-discrimination) in terms of estimated residual (at a significant χ^2 probability). Yet, the overall item-trait interaction χ^2 value was 178 (115 degrees of freedom, p<0.001), indicating that the items were not working as expected across different levels (i.e., class intervals) of HRQoL in the sample.

The mean local item dependency in the PedsQL 4.0 GCS was estimated at −0.04. In total, 13% (34 of 253) of all item pairs exhibited a mean residual correlation >0.20 above the mean correlation for all items, and 8% (21 of 253) >0.30. The issue was particularly prominent for item 5 ("Taking a bath or shower by himself") and item 6 ("Doing chores around the house") (mean correlation 0.60), item 2 ("Running") and item 3 ("Participating in sports activity or exercise") (mean correlation 0.52), and item 19 ("Paying attention in class") and item 21 ("Keeping up with schoolwork") (mean correlation 0.50). Additional item dependency results are available as supplemental material (online).

The PSI was estimated at 0.903. By investigating residual principal component loadings, the PedsQL 4.0 GCS was found to be multidimensional, with 24% statistically significant t-tests (p<0.05) (95% confidence interval: 21%−26%). The second factor (i.e., the primary contributor to the variance of the data with the Rasch factor discounted) particularly influenced the first eight questions (i.e., the Physical Functioning domain of the instrument).

3.2 Item Thresholds

Disordered thresholds were identified for 87% (20 of 23) of all items (Figure 2). This indicates that participants generally had difficulty discriminating between response categories given their level of HRQoL. For most items, issues with level thresholds concerned "Never" versus "Almost Never" and/or "Often" versus "Nearly Always". The distribution of item threshold locations on the estimated continuum is presented in Figure 3. Overlapping item thresholds (i.e., that several thresholds measure the same level of HRQoL) was relatively common, in particular at 0.0 logits (covered by 12 individual item thresholds) and −0.2 logits (covered by 9 thresholds). Evident from the figure, there were also issues with item threshold targeting of participants with relatively high ability levels (2 logits).

3.3 Person Fit to the Rasch Model

The distribution of participants on the estimated continuum is presented in Figure 3. The mean location of individual responses was 0.453 (range: −0.854 to 0.992), indicating that the sample exhibited a higher level of HRQoL than what would be expected on average from

the included items, with a mean fit residual of −0.255 (SD: 1.504). There were no floor or ceiling effects (i.e., no extreme values).

3.4 Differential Item Functioning

Analysis of scale stability showed that there was no significant uniform differential item functioning (i.e., a systematic difference across the full range of level of HRQoL) or non-uniform differential item functioning (i.e., non-uniformity in the differences across level of HRQoL) by glucocorticoid use (any lifetime exposure vs. none; $p>0.002$ and $p>0.005$, respectively). In contrast, the scale exhibited significant uniform differential item functioning by ambulatory stage (early ambulatory vs. late ambulatory vs. non-ambulatory) for 30% (7 of 23) of items in the PedsQL 4.0 GCS (non-uniform differential item functioning was not detected, $p>0.001$). Additional differential item functioning results are available as supplemental material (online).

3.5 Sensitivity Analysis

Results from the Rasch analysis of the English-language version of the PedsQL 4.0 GCS, comprising of a total of 90 patients from Australia and Canada, were similar to those derived for the total sample population, with a few exceptions. Most notably, in the ranking of items in terms of difficulty, "Missing school to go to the doctor or hospital" and "Having hurts or aches" were indicative of a relatively greater loss in HRQoL. Moreover, item 14 ("Getting along with other children") no longer displayed model misfit, and the overall item-trait interaction χ^2 value was 39 (23 degrees of freedom, $p=0.022$). Yet, the mean local item dependency was almost identical (−0.0421 vs. −0.0424), disordered thresholds were identified for 70% (16 of 23) of all items, and the English-language scale was found to be multidimensional (19% statistically significant t-tests, 95% CI: 14%−23%). Additional sensitivity analysis results are available as supplemental material (online).

4 Discussion

The objective of this study was to examine the psychometric properties of the PedsQL 4.0 GCS administered to patients with DMD recruited as part of the DNHS, the largest prospective multicentre study to date in this disease population. Taken together, the results from our assessment indicate that the scale may not be fit for purpose to measure HRQoL in DMD. A discussion of our specific findings follow below.

The Rasch analysis revealed that the most difficult PedsQL 4.0 GCS items were those associated with emotional well-being (e.g., being teased by other children, feeling sad, and not making friends). In contrast, the least difficult items were those mainly reflecting physical disability (e.g., lifting heavy objects, participating in sports, and running). These findings – which should be helpful to inform the development of new scales measuring HRQoL in DMD and similar illnesses – suggest that the largest loss in HRQoL was captured by items associated with morbidity and disability in addition to and beyond the primary somatic manifestation of DMD. Similar rankings of items have been reported as part of previous Rasch analyses of the PedsQL 4.0 GCS, for example among Canadian preschool children with refractive errors.[8] A comparable pattern was also identified as part of a

psychometric analysis of the 3.0 Neuromuscular Module to the PedsQL scale in patients with DMD from the UK and the US.[10] One potential explanation for these results could be related to coping mechanisms (i.e., the process of adapting to a changed health state and of accommodating illness[27]), as previously described for this disease population.[28] However, further research into these topics is warranted. The relative importance of emotional and social functioning among children and adolescents with DMD is also of relevance for the clinical and social management of the disease, including appropriate school support to help maintain and promote HRQoL in the presence of progressive physical disability.

In accordance with previous research,[7,8] the majority of all items in the PedsQL 4.0 GCS were found to have disordered thresholds. This means that patients generally had difficulty discriminating between response categories given their HRQoL (i.e., ability level) as estimated by the scale. Similar properties were also found in a Rasch analysis of the PedsQL 3.0 Neuromuscular Module in DMD.[10] Put differently, these findings suggest that patients (and caregivers) perceive the current level structure of the PedsQL 4.0 GCS as ambiguous, which may be the result of, for example, too many response categories, vague instructions for completing the scale (e.g., suboptimal descriptions and exemplifications of response categories and items), and/or unclear labels.

On average across scale items, 72% of patients reported having any problems (range: 52% −91%) and 51% having problems at least sometimes (range: 26%–81%), indicating that the items were clinically relevant to this disease population. No patients indicated maximum problems (i.e., the "Almost Always" item-level) nor minimum problems (i.e., the "Never" item-level) across all scale items. Accordingly, there were no minimum or maximum scores (i.e., non-informative extreme scores) in our sample. That being said, on average, patients in our sample had a higher level of HRQoL than what would be expected on average from the included items. Some might find this surprising, given that the PedsQL 4.0 GCS was developed as a generic instrument, relevant and applicable also to healthy populations. Indeed, despite that 62% (204 of 329) of the study cohort were in the late-ambulatory disease stage (with a stand from supine time of >5 seconds) or non-ambulatory stage, <27% (88 of 329) of patients were estimated to have relatively low HRQoL as indicated by negative logit location on the scale. Similar findings were reported by Amin et al.[7] in their analysis of the PedsQL 4.0 GCS administered to Canadian preschool children with refractive errors. One potential explanation of this result in the context of DMD could be related to the fact that many patients are likely to rely upon medical devices and aids to perform their day-to-day tasks and activities as captured by the PedsQL, which would thus generate estimates of a relatively higher person ability and/or lower item difficulty. Another potential explanation is that from a quality of life standpoint, the impact of even severe limitations in strength and mobility may be moderated by other supportive factors, such as successful peer and family relationships, or accessible educational and community environs.

We found items within the PedsQL GCS to be heavily locally dependent, with almost half exhibiting a residual correlation >0.2 above the mean correlation with at least one other item. In addition, several items displayed residual correlations with 3–5 other items. Considering that a few items within the scale appear to be measuring closely related aspects of life using similar wording, this result may not come as a surprise. For example, our

analysis show that the added value (in the context of measuring HRQoL in patients with DMD) of assessing ability to independently take a bath or shower *and* do chores around the house, or ability to run *and* to participate in sports activity or exercise may be limited. Interestingly, our analysis of individual item fit to the Rasch model did not identify any redundant items as indicated by a large negative fit residual, although problems were noted for item 6 ("Doing chores around the house").

Overall, the PedsQL 4.0 GCS did not demonstrate good targeting in our cohort as illustrated by the lack of overlap of person ability and item difficulty in the person-item threshold map (Figure 3). This was evident at both lower (<−2.0 logits) as well as higher (2.0 logits) ability levels on the estimated continuum. In addition, we found non-trivial issues with overlapping thresholds, which means that several items were duplicating the capacity to discriminate at that level of ability.

Estimates of PSI indicated good reliability of the PedsQL 4.0 GCS in DMD. However, it is well-known that issues with, for example, local dependency may cause spuriously inflated reliability. Accordingly, although above the proposed minimum threshold value of 0.80,[29] until further evidence is available, we recommend that this index should be interpreted and compared with caution. We also found evidence of significant differential item functioning by disease stage, which means that the scale does not behave similarly across the progressions sequence in DMD (in our case defined in terms of ambulatory status, time to stand from supine, and/or predicted forced vital capacity).

Lastly, our psychometric examination showed that the PedsQL 4.0 GCS may not be regarded a unidimensional, interval rating-scale of HRQoL among patients with DMD. This means that the scale fails to adhere to the epistemological requirements for stable, objective measures of social variables.[15] Accordingly, as an ordinal measure, the PedsQL 4.0 GCS does not allow for basic arithmetic operations, including calculation of mean scores, or changes in mean total scores. These data also suggest that it is not meaningful to compare mean total scale scores across trials, studies, or samples, or even individual scores between patients, as they are not invariant. However, it is important to keep in mind that these results concern the total scale score; accordingly, the PedsQL 4.0 GCS domain scores might still be interval (although still subject to limitations concerning disordered thresholds, etc.). Finally, it should be noted that it may be possible to improve the properties of the PedsQL 4.0 GCS using, for example, testlets and re-scoring of individual item thresholds, but this is an extensive undertaking beyond the scope of the current study.

Strengths of our study include a comparatively large sample of patients with DMD, a comprehensive psychometric assessment using a modern methodology anchored in the theory of fundamental measurement, and formal adjustment for multiple comparisons. The main limitation of our work concerns external validity, as it is not known to what degree included patients are representative of the total DMD population. That being said, the collected clinical and epidemiologic data were characteristic for the different patient groups (as defined by age and/or clinical disease milestones), which suggest that the discrepancy between the sample and study population is limited. Finally, despite our sizable cohort, we were unable to fully explore some psychometric properties of the PedsQL 4.0 GCS,

for example scale stability for patients with advanced disease, different age-formats of the instrument, as well as replies to all included language-versions of the scale. Yet, similar psychometric issues were found for the English-language sample as for the combined cohort, which indicates that some of the adverse properties of the PedsQL 4.0 GCS may not primarily be related to linguistic features, but rather the item-level structure of the scale continuum.

In summary, our Rasch analysis of the PedsQL 4.0 GCS administered to children and adolescents with DMD revealed significant psychometric issues, including local item dependency, disordered thresholds, suboptimal targeting, poor item-trait interaction, differential item functioning, and multidimensionality. Accordingly, based on our analysis, it appears as if the PedsQL 4.0 GCS total score fails to successfully operationalize a quantitative conceptualization of HRQoL in patients with DMD and should be used with caution in this indication until further evidence is made available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Emery AE. The muscular dystrophies. Lancet 2002; 359: 687–695. [PubMed: 11879882]

2. Verhaart IEC, Aartsma-Rus A. Therapeutic developments for Duchenne muscular dystrophy. Nat Rev Neurol 2019; 15(7): 373–386. [PubMed: 31147635]

3. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol 2007; 6: 1094–1105. [PubMed: 18031706]

4. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. Med Care 2001; 39(8): 800–812. [PubMed: 11468499]

5. Davis SE, Hynan LS, Limbers CA, et al. The PedsQL in pediatric patients with Duchenne muscular dystrophy: feasibility, reliability, and validity of the Pediatric Quality of Life Inventory Neuromuscular Module and Generic Core Scales. J Clin Neuromuscul Dis 2010; 11: 97–109. [PubMed: 20215981]

6. Lim Y, Velozo C, Bendixen RM. The level of agreement between child self-reports and parent proxy-reports of health-related quality of life in boys with Duchenne muscular dystrophy. Qual Life Res 2014; 23(7): 1945–52. [PubMed: 24566887]

7. Amin L, Rosenbaum P, Barr R, et al. Rasch analysis of the PedsQL: an increased understanding of the properties of a rating scale. J Clin Epidemiol 2012; 65(10): 1117–1123. [PubMed: 22910540]

8. Lamoureux EL, Marella M, Chang B, et al. Is the pediatric quality of life inventory valid for use in preschool children with refractive errors? Optom Vis Sci 2010; 87(11): 813–822. [PubMed: 20852452]

9. Kook SH, Varni JW. Validation of the Korean version of the pediatric quality of life inventory 4.0 (PedsQL) generic core scales in school children and adolescents using the Rasch model. Health Qual Life Outcomes 2008; 6: 41. [PubMed: 18518951]

10. Landfeldt E, Mayhew A, Straub V, Lochmüller H, Bushby K, Lindgren P. Psychometric analysis of the Pediatric quality of life inventory 3.0 neuromuscular module administered to patients with Duchenne muscular dystrophy: a Rasch analysis. Muscle Nerve 2018; 58(3): 367–373. [PubMed: 29466827]

11. McDonald CM1, Henricson EK, Abresch RT. The cooperative international neuromuscular research group Duchenne natural history study--a longitudinal investigation in the era of

glucocorticoid therapy: design of protocol and the methods used. Muscle Nerve 2013; 48(1): 32–54. [PubMed: 23677550]

12. 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(10119): 451–461. [PubMed: 29174484]

13. Henricson EK, Abresch RT, Cnaan A, et al. The cooperative international neuromuscular research group Duchenne natural history study: glucocorticoid treatment preserves clinically meaningful functional milestones and reduces rate of disease progression as measured by manual muscle testing and other commonly used clinical trial outcome measures. Muscle Nerve 2013; 48(1): 55–67. [PubMed: 23649481]

14. Rasch G. Probabilistic models for some intelligence and attainment tests (1st Edition). Copenhagen: Danish Institute for Education Research; 1960.

15. Wright B. A history of social science and measurement. Educ Meas 1997; 52: 33–52.

16. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009; 13.

17. da Rocha NS, Chachamovich E, de Almeida Fleck MP, Tennant A. An introduction to Rasch analysis for Psychiatric practice and research. J Psychiatr Res 2013; 47(2): 141–148. [PubMed: 23069651]

18. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978; 43, 561–73.

19. Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? Med Care 2004; 42: I7–I16. [PubMed: 14707751]

20. Lundgren Nilsson Å, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM™) revisited. J Rehabil Med 2011; 43(10): 884–891. [PubMed: 21947180]

21. Tennant, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007; 57(8): 1358–1362. [PubMed: 18050173]

22. Masters G. A Rasch model for partial credit scoring. Psychometrika 1982; 47: 149–174.

23. Fisher W. Reliability Statistics. Rasch Measurement Transactions 1992; 6(3): 238.

24. Tennant A, Penta M, Tesio L, et al. Assessing and adjusting for cross cultural validity of impairment and activity limitation scales through Differential Item Functioning within the framework of the Rasch model: the Pro-ESOR project. Medical Care 2004; 42: 37–48.

25. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3(2): 205–231. [PubMed: 12011501]

26. Tennant A, Pallant JF. Unidimensionality matters. Rasch Meas Trans 2006; 20: 1048–1051.

27. Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. Health Technol Assess 2001; 5(4).

28. Landfeldt E, Lindgren P, Bell C, et al. Health-related quality of life in patients with Duchenne muscular dystrophy: a multinational, cross-sectional study. Dev Med Child Neurol 2016; 58(5): 508–515. [PubMed: 26483095]

29. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009; 13.

Highlights

- The PedsQL 4.0 GCS is a commonly used measure of health-related quality of life (HRQoL) in patients with Duchenne muscular dystrophy (DMD).

- We evaluate the psychometric properties of the PedsQL 4.0 GCS using Rasch analysis.

- We show that the PedsQL 4.0 GCS may not be a valid measure of HRQoL in DMD.
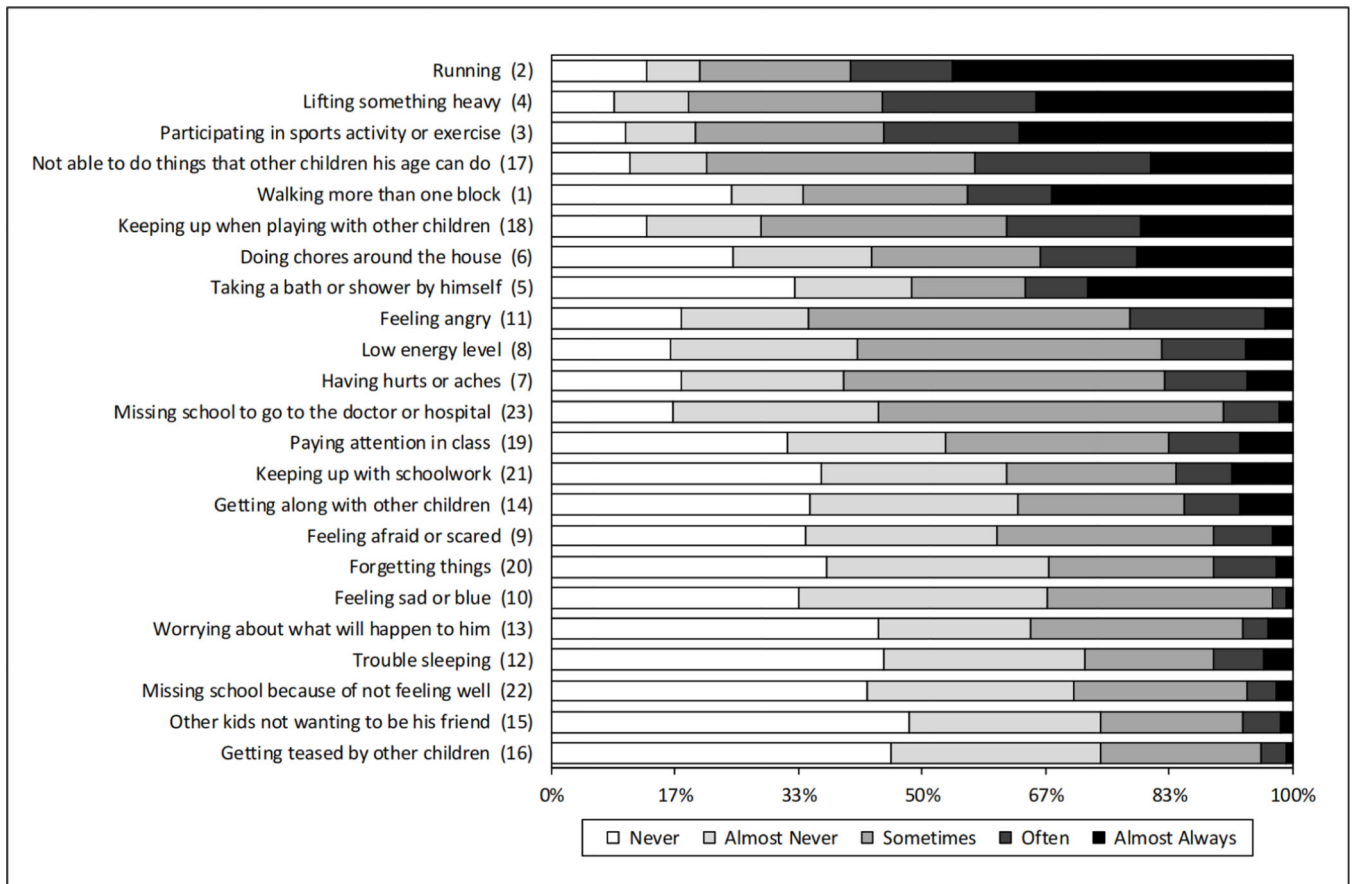
Figure 1: Distribution of replies to the PedsQL 4.0 GCS

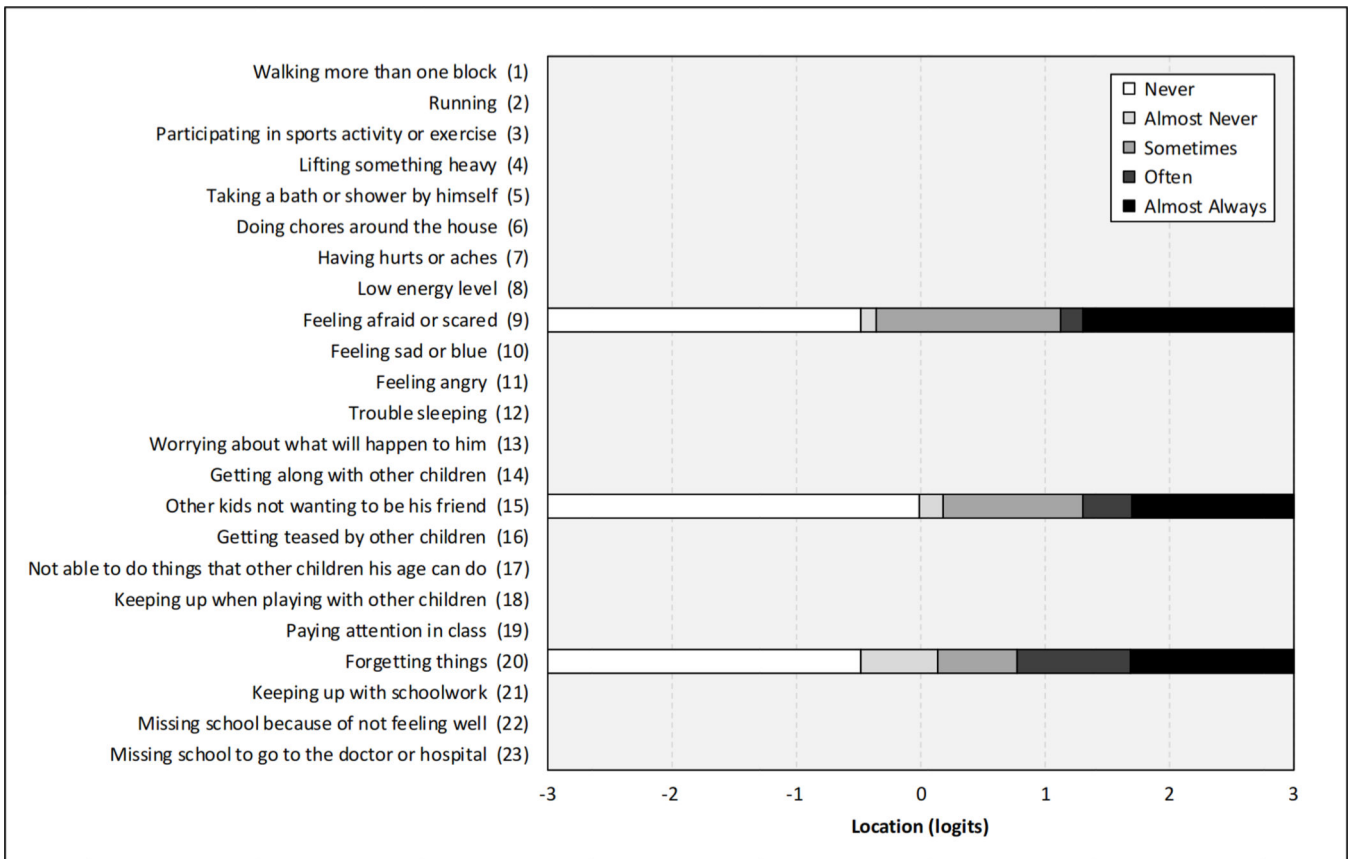Note: The number in parenthesis represent the item number in the PedsQL 4.0 GCS.

Figure 2: Threshold map

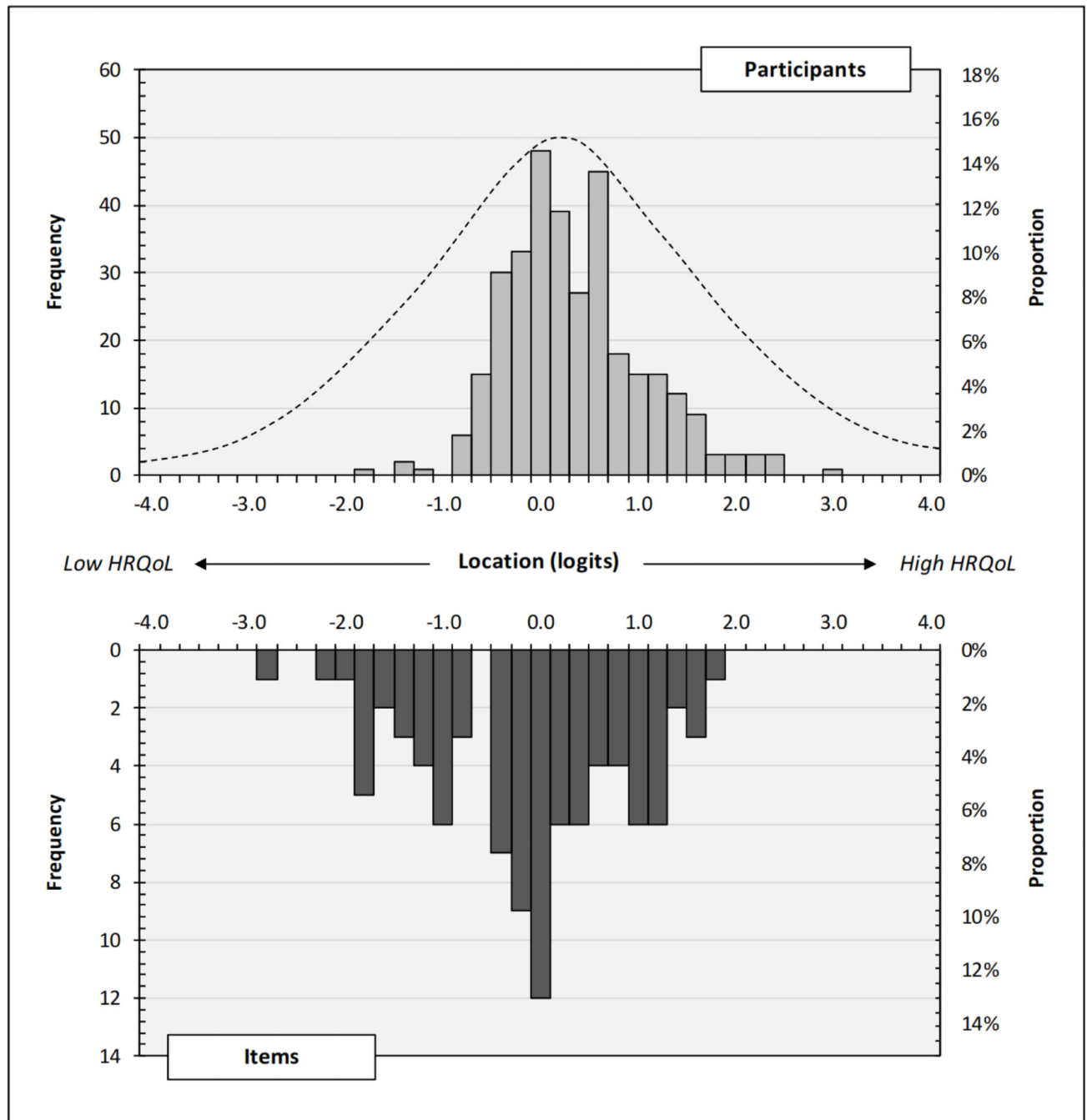Note: Missing bars represent disordered thresholds.

Figure 3: Targeting map
Note: The top chart area shows the location of the participants (n=329) on the interval logit scale representing level of HRQoL (a low number represents low HRQoL, and vice versa). The dotted line represents the information curve. The bottom chart area shows the location of the PedsQL 4.0 GCS item thresholds (23×4=92) on the same logit scale (a low number represents low item difficulty, and vice versa).

Table 1:

Demographic and clinical characteristics of the study sample

Sex, male	329 (100%)
Age, mean (SD) [range] years	9 (4) [3–18]
Region/country [a]	
Argentina	15 (5%)
Australia	35 (11%)
Canada	55 (17%)
Europe (Israel, Italy, and Sweden)	42 (13%)
India	49 (15%)
The US[b]	133 (40%)
Race	
Caucasian	230 (70%)
Black	4 (1 %)
Asian	63 (19%)
Other[c]	32 (10%)
Disease stage	
Early ambulatory	112 (34%)
Late ambulatory[d]	134 (41%)
Early non-ambulatory	70 (21%)
Late non-ambulatory[e]	13 (4%)
Lifetime exposure to glucocorticoids	249 (76%)

Note: Data presented as n (%) if not specified otherwise. Because of rounding, percentages may not add up to exactly 100%.

[a]Data not available for all countries separately.

[b]Includes Puerto Rico.

[c]Includes Pacific Islander and Native American.

[d]Time to stand from supine >5 seconds.

[e]Predicted forced vital capacity <30%.

Table 2:

Individual item fit to the Rasch model

Item (item number)	Item location*	SE	Fit residual (observed-expected)	x^2	x^2 probability
Getting teased by other children (16)	−0.85	0.07	0.93	2.91	0.7139
Feeling sad or blue (10)	−0.81	0.07	−1.09	15.73	0.0076
Other kids not wanting to be his friend (15)	−0.79	0.06	−1.08	5.43	0.3656
Missing school because of not feeling well (22)	−0.65	0.06	0.79	7.48	0.1875
Forgetting things (20)	−0.52	0.06	0.74	4.23	0.5173
Worrying about what will happen to him (13)	−0.47	0.06	−0.58	9.98	0.0758
Trouble sleeping (12)	−0.43	0.06	1.00	7.60	0.1795
Feeling afraid or scared (9)	−0.42	0.06	1.18	1.66	0.8944
Missing school to go to the doctor or hospital (23)	−0.23	0.07	1.45	10.38	0.0653
Keeping up with schoolwork (21)	−0.12	0.05	−0.31	2.37	0.7955
Getting along with other children (14)	−0.12	0.05	2.80 †	21.92	0.0005 †
Paying attention in class (19)	−0.08	0.05	2.47	5.80	0.3259
Feeling angry (11)	−0.03	0.06	2.46	7.65	0.1764
Having hurts or aches (7)	0.07	0.06	1.41	5.50	0.3585
Low energy level (8)	0.10	0.06	−0.28	9.43	0.0932
Taking a bath or shower by himself (5)	0.32	0.04	−0.51	1.10	0.9537
Doing chores around the house (6)	0.34	0.05	−2.38	11.66	0.0397
Walking more than one block (1)	0.53	0.05	−1.14	6.26	0.2821
Keeping up when playing with other children (18)	0.61	0.05	1.48	14.12	0.0149
Not able to do things that other children his age can do (17)	0.69	0.06	−0.21	2.11	0.8336
Running (2)	0.93	0.05	0.27	9.19	0.1018
Participating in sports activity or exercise (3)	0.94	0.05	−0.81	5.74	0.3322
Lifting something heavy (4)	0.99	0.05	−1.67	10.23	0.0690

Note: Mean item fit residual: 0.301 (SD: 1.385).

* A low number represents high difficulty (i.e., low HRQoL as expressed by the item), and vice versa.

† Denotes a misfitting item (i.e., fit residual >|2.5|) or χ^2 probability<0.002174).