



This is a repository copy of *Measuring quality of life in Duchenne muscular dystrophy : a systematic review of the content and structural validity of commonly used instruments.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/163997/>

Version: Published Version

Article:

Powell, P. orcid.org/0000-0003-1169-3431, Carlton, J. orcid.org/0000-0002-9373-7663, Woods, H.B. orcid.org/0000-0002-0653-9803 et al. (1 more author) (2020) Measuring quality of life in Duchenne muscular dystrophy : a systematic review of the content and structural validity of commonly used instruments. *Health and Quality of Life Outcomes*, 18. 263.

<https://doi.org/10.1186/s12955-020-01511-z>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

REVIEW

Open Access



Measuring quality of life in Duchenne muscular dystrophy: a systematic review of the content and structural validity of commonly used instruments

Philip A. Powell^{1,2*} , Jill Carlton¹, Helen Buckley Woods¹ and Paolo Mazzone¹

Abstract

Duchenne muscular dystrophy (DMD) is an inherited X-linked neuromuscular disorder. A number of questionnaires are available to assess quality of life in DMD, but there are concerns about their validity. This systematic review aimed to appraise critically the content and structural validity of quality of life instruments for DMD. Five databases (EMBASE, MEDLINE, CINAHL, PsycINFO, and Cochrane Library) were searched, with supplementary searches in Google Scholar. We included articles with evidence on the content and/or structural validity of quality of life instruments in DMD, and/or instrument development. Evidence was evaluated against the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) criteria. Fifty five articles featured a questionnaire assessing quality of life in DMD. Forty instruments were extracted and 26 underwent assessment. Forty-one articles contained evidence on content or structural validity (including 37 development papers). Most instruments demonstrated low quality evidence and unsatisfactory or inconsistent validity in DMD, with the majority not featuring direct validation studies in this population. Only KIDSCREEN received an adequate rating for instrument design and a satisfactory result for content validity based on its development, yet, like the majority of PROMs, the measure has not been directly validated for use in DMD. Further research is needed on the validity of quality of life instruments in DMD, including content and structural validity studies in this population.

Keywords: Content validity, Duchenne muscular dystrophy, Patient reported outcome measures, Quality of life, Structural validity

Introduction

Duchenne muscular dystrophy (DMD) is an X-linked neuromuscular disorder with an estimated incidence of 1 in 3802–6291 live male births [1, 2]. The disease causes progressive muscle weakness due to an absence of the dystrophin protein, which functions to help keep muscle cells intact. Diagnostic symptoms and functional

impairment are evident from as early as two years old and average life expectancy of people with DMD is approximately 25 years [3], although increasingly people with DMD are surviving into their fourth and even fifth decades [4]. The disease progresses through four recognised clinical stages characterised by increased muscle weakness, impaired ambulation and motor functioning, and cardiovascular and respiratory problems [5]. There is no cure for the disease. Current clinical efforts are focused on slowing disease progression and improving the health-related quality of life (QoL) of people with DMD,

* Correspondence: p.a.powell@sheffield.ac.uk

¹School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

²Department of Economics, University of Sheffield, 9 Mappin Street, Sheffield S1 4DT, UK



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and health interventions are necessarily evaluated for their cost effectiveness against this objective.

In order to attempt to measure QoL in people with DMD a number of both generic (such as the EQ-5D [6, 7]) and condition-specific (such as the MDCHILD [8]) patient reported outcome measures (PROMs) are used. However, concerns have been raised about the validity of existing PROMs to comprehensively assess QoL in DMD [9]. Given that a number of generic and condition-specific questionnaires are available, researchers and clinicians have to make a critical choice on which measure may be most appropriate for assessing QoL in people with DMD. In order to help inform this decision, evidence-based guidance is needed on the relative validity and psychometric performance of these instruments. There are a number of reviews exploring QoL and associated measures in DMD, with some providing very basic information on their psychometric properties [9, 10]. However, no reviews to date have appropriately evaluated the content validity of available measures when it comes to assessing QoL in DMD. This is a striking omission; content validity has been defined by the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) group as the most important property of a PROM [11–13]. Furthermore, prior reviews on QoL in neuromuscular disorders have either not referred to, or used an outdated version of, COSMIN guidance. In the current review we used up-to-date COSMIN methodology to assess the content and structural validity of QoL PROMs in DMD [11].

Content validity refers to the extent that the content of a PROM adequately reflects the target construct that is intended to be measured [14]. It can be subdivided into the judged ‘relevance’, ‘comprehensiveness’, and ‘comprehensibility’ of a PROM, in assessing the construct of interest within a target population and context [13]. ‘Relevance’ of a PROM refers to whether the items are relevant for the construct, target population, and context of use of interest; the response options and recall period of a PROM should also be appropriate and relevant. ‘Comprehensiveness’ is used to describe the extent to which all key aspects of the construct of interest are covered in the PROM. Finally, ‘comprehensibility’ pertains to the understanding of the items and response options by the population of interest [13].

A thorough assessment of a PROM’s content validity should include studies presenting information on content validity in the population of interest, but also consider the initial PROM development paper(s) (i.e. literature describing studies on the development of the PROM) and the content of the PROM itself [12, 13]. The consideration of development studies is important, because the quality of how the PROM was developed

(e.g. was there a clear description of the construct to be measured? were patients involved? etc.) necessarily has an impact on the evaluation of the content validity of a PROM in its subsequent use. Thus, COSMIN recommends that the quality of PROM development is rated and assessed prior to the quality of any content validity studies [13]. Furthermore, content validity should form the first step of the assessment of the validity of a PROM, as it is integral to that PROM’s usefulness in doing the job it was designed to do, and influences all other measurement properties [15, 16]. For example, a psychometrically responsive and internally consistent instrument is of little use if it is not measuring what it is intended to measure.

COSMIN guidance states that the second most important form of the validity assessment of a PROM is structural validity [15, 16]. Structural validity describes the extent that scores derived from a measure adequately reflect the dimensionality of the construct being measured [17]. Quality of life is usually defined, and thus measured, as a multidimensional construct. Therefore, PROMs that feature multiple dimensions of QoL should be assessed to check they accurately represent the multidimensional structure of QoL in the population of interest. If PROMs are designed to target a single dimension of QoL, assessments should be undertaken to empirically demonstrate their unidimensional nature in the target population. If such assessments are not undertaken, subsequent interpretation of the data (e.g. through generating dimensional scores) may be inaccurate. For the purposes of this review, we define QoL as a multidimensional construct involving physical (e.g., pain, fatigue), psychological (e.g., mood, self-efficacy), and social (e.g., participation, stigma) components, based on the Comprehensive Model of QoL in Muscular Dystrophy (CMQM) [9], and use this to define the construct of interest. We choose to define QoL as a subjective construct and do not include purely functional performance or assessment scales that may impact on QoL. In this review, we consider multi-item PROMs that assess *at least* one aspect of QoL in people with DMD.

When evaluating a PROM, content and structural validity can be meaningfully assessed against up-to-date published standards by the COSMIN group, derived from international expert consensus [12, 14]. These ratings incorporate actual evidence on PROM validity and the quality of that evidence. For example, regarding a PROM’s ‘comprehensiveness’, a positive rating can be given based on a content validity study if: (i) the study quality was not rated as inadequate; (ii) patients or professionals were interviewed; and (iii) no key concepts were missing. For structural validity, a positive rating is given if good model fit is observed in CFA or in IRT/Rasch (see Methods), and can be appraised alongside a

rating for the study's quality. The full COSMIN standards and methodology for assessing PROMs are comprehensive and available in accompanying guidance manuals [13, 16], which were adhered to when conducting the current review.

This systematic review has been designed to evaluate the content and structural validity of QoL measures used in people with DMD using updated COSMIN guidance [13, 16], to provide researchers and clinicians with a robust evidence-base to help them when selecting PROMs to measure QoL in the Duchenne population. The review makes a unique contribution to the literature in being the first to assess the content validity of PROMs used in DMD and to apply an up-to-date and thorough COSMIN assessment of these measures. There are two main questions being addressed:

- 1) Which PROMs have been used to assess QoL in published research with boys and men diagnosed with DMD?
- 2) What is the content and structural validity of these PROMs for use in assessing QoL in boys and men with DMD?

Methods

The review protocol was registered on PROSPERO [18]. This systematic review has been reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist [19].

Search strategy and selection criteria

This review contains two searches. The first search (Search A) was designed to identify PROMs used to measure QoL in DMD in peer-reviewed publications. The second search (Search B) was used to identify literature reporting on the measurement properties of these PROMs in DMD. Search B also included the recommended practice of searching for the development papers of PROMs to enable a full COSMIN assessment of their content validity [12, 13, 15]. Full copies of the searches are contained in Additional file 1, for reproducibility.

Search A and selection criteria

Search A was conducted on 11th April 2018, searching EMBASE, MEDLINE, CINAHL, PsycINFO, and the Cochrane library, from inception. No restrictions on date or language were applied to the search. Search A terms are illustrated in full in Additional file 1 and included: (I) Duchenne muscular dystrophy (and Duchenne*) AND ((II) a search filter provided by the PROM group at the University of Oxford to identify PROMs (available online [20] and in Additional file 1)

OR (III) PROMs known to be used in people with DMD based on a prior rapid review of the literature [21].

The following selection criteria were applied to the results of Search A by two independent reviewers: (I) published in English as a full-text original research article (i.e. not including abstracts, editorials, or reviews); (II) used a self-reported, multi-item PROM to assess *at least one* aspect of QoL in males diagnosed with DMD (assisted or proxy-reported versions of PROMs were considered for inclusion so long as a self-report version of that PROM exists); and (III) in case of studies involving mixed clinical samples, at least 75% of the sample (or subgroup), on which data from the PROM was reported, was male diagnosed with DMD. The inclusion criteria were first applied to titles and abstracts of the hits from Search A. Records were selected for full-text review if they matched the selection criteria, potentially matched the criteria, or if doubt existed. Any discrepancy was resolved by a third reviewer. Full text articles were then screened for selection using the selection criteria by two independent reviewers. Any disagreements were resolved by a third reviewer through discussion. Finally, the PROMs themselves identified in the articles were reviewed by two independent reviewers to ensure they met the requisite inclusion criteria (i.e. assessing an aspect of QoL).

Search B and selection criteria

Search B was conducted on 18th September 2018, with initial searches on EMBASE, MEDLINE, CINAHL, PsycINFO, and the Cochrane library, from inception. No restrictions on date or language were applied to the search. Search B terms are illustrated in full in Additional file 1 and included: term (I) from Search A AND ((II) PROMs identified in Search A OR term (III) from Search A) AND (IV) a search filter¹ by the COSMIN group for identifying studies on measurement properties [22] (available online [23] and in Additional file 1). Over and above that of Search A, the following additional selection criteria was applied to the results of Search B: (IV) described data on the content and/or structural validity of the PROMs identified in Search A in males diagnosed with DMD; (V) included a PROM validated in English, with a free/review copy available to access.

As recommended in the COSMIN approach [15], follow up searches were conducted on Google Scholar to identify key development papers for the PROMs identified in Search A and taken forward for review (see Section 3.2). Google Scholar was searched (last searched

¹A less-restrictive and simplified version of this filter was applied to the Cochrane Library and CINAHL searches, due to a low number of hits prior to filtering for the measurement properties (see Additional file 1 for full search strategies).

14th November 2018) with the names and acronyms of the PROMs (version numbers omitted) and the first 100 hits were screened for inclusion [15]. Search results were initially screened by title, with any relevant and potentially relevant papers exported to a database. Following the removal of duplicates against the primary searches, records were screened by abstract and then full text against selection criteria. As per COSMIN guidance, development papers for the PROMs were not subject to any of the inclusion criteria noted above and were included in any published form [13, 15]. Results of the searches were screened for inclusion by two reviewers.

Finally, citation tracking of all eligible articles identified in Search B was conducted by reviewing references and citations on Google Scholar (last searched 6th February 2019) for any articles not identified in the initial searches that may meet the inclusion criteria. All references and citations were reviewed, except where citations became unmanageable (i.e. > 500 citations), when “Duchenne” was searched for within the citing articles to filter the hits for manageable review. Search results were initially screened by title, with any relevant and potentially relevant papers exported to a database. Following the removal of duplicates against the primary searches, records were screened by abstract and then full text against selection criteria. Results of the searches were screened for inclusion by two reviewers.

Data extraction and COSMIN risk of bias assessment

Data extraction was undertaken by two reviewers using a pre-prepared data extraction sheet, with consensus on any ambiguities reached through discussion. The data extraction sheet was informed by the tools developed by COSMIN on reporting guidance [16], and included study characteristics (authors, year, journal, country, language, study type); details of the PROM used (name, mode of administration, recall period, total N subdomains, subdomain names, total and subdomain N items, total and subdomain response levels, total and subdomain score ranges); DMD sample characteristics if applicable (N , age, percentage ambulatory, total and subdomain PROM score, total and subdomain observed ranges); details of PROM development if applicable (construct definition, target population, original language, intended context of use, patient involvement); details of content validity results if applicable (summarised results, e.g. findings from a cognitive debriefing exercise); and details of structural validity results if applicable (analytic model, summarised results, e.g. fit statistics, tests of model assumptions for IRT/Rasch).

The methodological quality of the PROM development papers, and studies on content and structural validity were assessed (at the study level) using up-to-date COSMIN standards via the new COSMIN risk of bias

checklist [24]. A total rating for relevance, comprehensiveness, and comprehensibility (content validity aspects) of a PROM is determined separately, alongside a total rating for the methodological quality of a structural validity study [24]. When rating the methodological quality of the studies, each COSMIN standard (or item) is ranked on a 4-point scale: “very good”, “adequate”, “doubtful”, and “inadequate”. Total ratings are determined using the lowest rating for any item for that study (i.e. worst score counts) [25]. Studies were initially rated independently by two reviewers, and, in the case of divergence, consensus was reached in a subsequent face-to-face meeting. This information on risk of bias is used to inform quality of evidence (see section 2.2).

Assessment of measurement properties

In order to synthesise and assess evidence on content validity, two reviewers independently rated the results of PROM development studies, content validity studies, and the content of the PROM itself on 10 COSMIN criteria [13], agreed upon by international consensus [12]. These criteria included: whether the included items were relevant for (I) the construct of interest, (II) the population of interest, and (III) the context of use of interest; whether the (IV) response options and (V) recall period were appropriate; whether (VI) all key concepts were included; whether (VII) the PROM instructions and (VIII) PROM items and response options were understood by the population of interest as intended; whether (IX) the PROM items were appropriately worded; and whether (X) the response options matched the question. Ratings for each source of evidence were made separately, using COSMIN guidance [13] (p.54) and could either be positive (+), negative (-), or indeterminate (?). Reviewers' ratings were made based on the judgement of the researchers, who have experience in PROM design and work with people with DMD, including direct qualitative research [26, 27]. When reviewers considered whether the items were relevant or comprehensive for the construct of interest, they were compared against the CMQM [9]. Accordingly, a PROM would be sufficiently comprehensive (+) if it included items covering physical, psychological, and social aspects of QoL. When judging the appropriateness of the recall period, reviewers considered any defined recall period of up to 4 weeks as appropriate (+), as children aged 8 years and above can recall up to this length of time with sufficient accuracy [28]. When rating the appropriateness of response options, bearing in mind the target sample (i.e. a child or adult PROM), reviewers took into account the numerical range, how the response options were visually displayed, and the perceived cognitive complexity of the options (including wordiness, degree of variation throughout the questionnaire, and the use of reversed ordering).

Following the above assessment, an overall (qualitatively synthesised) judgment on the relevance, comprehensiveness, and comprehensibility of each PROM was made, which could be sufficient (+), insufficient (-), or inconsistent (\pm), using COSMIN guidance [13] (p.58). For example, if all sources of evidence were rated positive (+) for relevance, then the overall rating for the PROM would be sufficient (+). As recommended by COSMIN [13], more weight was given to content validity studies, then development studies, then ratings of the PROM by reviewers. Ratings were compared and combined across the two reviewers by consensus. As per COSMIN guidance [13], only available evidence was taken into account when assessing content validity, so, for example, if there were no content validity studies in DMD available for that PROM, assessment was made based on the ratings of any PROM development studies and the ratings of reviewers. The fact that the PROM had no content validation studies in DMD is then reflected in a lower quality of evidence rating (see below). An example content validity rating spreadsheet for the KIDSCREEN-52, including the rules for synthesising the individual ratings is included in Additional file 2.

Evidence on structural validity was assessed against the updated COSMIN criteria for good measurement properties, using the same rating scale as above [16]. Specifically, a positive (+) rating would be given for a CFA featuring a CFI, TLI or comparable measure > 0.95 OR RMSEA < 0.06 OR SRMR < 0.08 . For an IRT/Rasch model, a positive (+) rating would be given for no violation of unidimensionally (e.g. assessed with the fit statistics above) AND no violation of local independence (e.g. residual correlations among items after controlling for the dominant factor < 0.20) AND no violation of monotonicity (e.g. evidenced graphically or item scalability > 0.30) AND adequate model fit (e.g. $\chi^2 < 0.01$, infit/outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized values > -2 and < 2). A negative (-) rating would be given if these criteria were not met in the data and an indeterminate (?) rating would be given if model fit was not reported.

Finally, the quality of the evidence was graded using a modified GRADE approach [29], as either “high”, “moderate”, “low”, or “very low”. The GRADE approach takes into account the risk of bias of studies (or study quality); (in) consistency across studies; imprecision (based on sample sizes); and indirectness (of evidence) [16]. The evidence is assumed to be high, then is downgraded by 1–3 points based on the degree of risk of bias (including quality and absence of content validity studies), 1–2 points based on inconsistency, and 1–2 points based on indirectness. Further details on how to apply all of the above criteria are provided elsewhere in comprehensive manuals, which were followed when conducting this

review [13, 16]. The quality of this systematic review itself was appraised against a recently developed COSMIN checklist to assess the quality of systematic reviews of health-related PROMs [22].

Results

Results of search A – PROMs used to measure quality of life in DMD

After removing duplicates, 1733 records were identified through database searching for Search A. Of these, 1521 were excluded at the title/abstract review stage, leaving 212 papers for full-text review. Of these 212 papers, 84 were excluded as they were not full-text published research articles; 25 did not meet the required sample criteria of at least 75% of the sample being boys or men with DMD; 21 were judged not to be assessing QoL; 16 were not published in English; and finally 11 papers did not feature a multi-item PROM. Five articles were additionally excluded during the review of the actual PROM used in the manuscript for not assessing QoL. Accordingly, a total of 50 records from the initial searches met the selection criteria for Search A. A further 5 articles that met the selection criteria for Search A were added as a result of citation tracking, giving a total of 55 records.

Table 1 summarises the PROMs used to assess QoL in DMD from the full-texts meeting the selection criteria at Stage 1 ($n = 55$). A total of 40 PROMs used to assess at least one aspect of QoL in DMD were identified in published research articles through database searching (the two HUI classification systems use the same 15-item PROM). The majority of the PROMs were multidimensional ($n = 32$), designed to assess a range of different facets of QoL. The remaining unidimensional scales were designed to assess: activity limitations (CALI); anxiety (GAD-7); depression (BDI, DIKJ, PHQ-9); fatigue severity (FSS); life satisfaction (SWLS); or quality of life/health-related quality of life unidimensionally (KIDSCREEN, SOLE). Twenty-four of the PROMs had versions designed for completion by adult or young adult respondents, and 26 had versions designed for children. The most popular PROMs used in published research articles assessing QoL in people with DMD were the PedsQL 4.0 GCS (18 articles); PedsQL 3.0 NMM (10 articles); and the SF-36 (8 articles).

Results of search B – evidence on measurement properties of PROMs

After removing duplicates, 92 records were identified through database searching for Search B. Of these, 51 had already been excluded during Search A. Eighteen unique records were found, 14 were excluded at title/abstract review stage, leaving 4 papers for full-text review. Of these 4 papers, 3 were excluded because they were

Table 1 Patient reported outcome measures assessing quality of life identified in published articles with samples of people with DMD

PROM	Respondent type	Recall Period	N dimensions (items)	Dimensions of quality of life assessed	Response options	Total score range	Origin	Validated English copy and development papers freely available for review?
36-item Short-Form Health Survey (SF-36) v1.0 [30–37]	Adult self-report	Varies by dimension	8 + a single item of perceived change in health (36 items)	Physical functioning; bodily pain; role limitations due to physical health problems; role limitations due to personal or emotional problems; emotional well-being; social functioning; energy/fatigue; general health perceptions	Varies by dimension	No total score calculated	USA	Yes
SF-36 Health Survey v2.0 [38]	Adult self-report	Varies by dimension	8 + a single item of perceived change in health (36 items)	Assumed same as SF-36 v1.0	Varies by dimension	No total score calculated	USA	No
Autoquestionnaire Qualité de vie Enfant Imagé (AUQEI) [39]	Child self-report	Unknown/undefined	4 + a total score (26 items)	Autonomy; leisure; functioning; family; total	0–3 rating scale (with pictures)	0–78 (raw)	France	No
Beck Depression Inventory I (BDI) [34]	Adult interview or self-report	Present/today	1 total score (21 items)	Depression	0–3 rating scale	0–63 (raw)	USA	Yes
Behavior Assessment System for Children (BASC) first edition [40]	Child self-report Parent report Teacher report	Unknown (withdrawn from use, superseded by BASC-II and BASC-III)	Unknown (withdrawn from use, superseded by BASC-II and BASC-III)	Unknown (withdrawn from use, superseded by BASC-II and BASC-III)	Unknown (withdrawn from use, superseded by BASC-II and BASC-III)	Unknown (withdrawn from use, superseded by BASC-II and BASC-III)	USA	No
Child Activity Limitations Interview (CALI) [40]	Child interview or self-report	Last 4 weeks	1 total score (8 items chosen from a set of 21)	Activity limitations	0–4 rating scale	0–32 (raw)	USA	Yes
Child Health Questionnaire - Parent Form 50 (CHQ-PF50) [41, 42]	Parent self-report	Last 4 weeks (past year for change in health)	14 (50 items)	Physical functioning; role/social limitations – physical; role/social limitations – emotional; role/social limitations – behavioral; general health perceptions; bodily pain/discomfort; family activities; parent impact – time; parent impact – emotion; self-esteem; mental health; behaviour; family cohesion; change in health	Varies by dimension	No total score calculated	USA	No

Table 1 Patient reported outcome measures assessing quality of life identified in published articles with samples of people with DMD (Continued)

PROM	Respondent type	Recall Period	N dimensions (items)	Dimensions of quality of life assessed	Response options	Total score range	Origin	Validated English copy and development papers freely available for review?
Children's Assessment of Participation and Enjoyment (CAPE) [43–45]	Child and young adults self-report or interview	Unknown/undefined	5 + a total score (55 items)	Diversity of activities; intensity of activities (frequency of participation); enjoyment of activities; with whom; where; total participation	Varies by dimension	0–55 (raw)	USA	No
Depressions-Inventar für Kinder und Jugendliche (DIKJ) 2nd edition [34]	Child self-report	Unknown/undefined	1 total score (26 items)	Depression	Unknown/undefined	0–46 (raw)	Germany	No
DISABKIDS generic module (DCGM-37) [34]	Child self-report Proxy report	Past 4 weeks	6 + a total score (37 items)	Independence; emotion; social inclusion; social exclusion; limitation; treatment; total	1–4 rating scale	37–148 (raw)	Multi-country	Yes
DISABKIDS – Smileys [34]	Child self-report Proxy report	Assumed same as DCGM-37	Assumed same as DCGM-37 (12 items)	Assumed same as DCGM-37	Assumed same as DCGM-37 (with pictures)	12–48 (raw)	Multi-country	No
Dutch Children AZL/TNO Questionnaire Quality of Life Short Form (DUC-25) [45]	Child self-report	Unknown/undefined	4 + a total score (25 items)	Physical; emotional; social; home functioning; total	0–4 rating scale	0–100 (raw)	Netherlands	No
EuroQoL 5-domain 3-Level (EQ-5D-3L) [46, 47]	Adult self-report Proxy report	Present/today	5 (5 items) + self-rated health VAS	Mobility; self-care; usual activities; pain/discomfort; anxiety/depression	1–3 rating scale	–0.594 – 1 (utility scores)	Multi-country	Yes
Fatigue Severity Scale (FSS) [36]	Adult self-report	Within last week	1 total score (9 items) + global fatigue VAS	Fatigue severity	1–7 rating scale	1–63 (raw)	USA	Yes
Generalized Anxiety Disorder Scale 7-item (GAD-7) [48]	Adult self-report	2 weeks	1 total score (7 items)	Anxiety	0–3 rating scale	0–21 (raw)	USA	Yes
Health Utilities Index Questionnaire mark 2 (HUI-2) 15Q [49, 50]	Child self-report Adult self-report Proxy report	During past 4 weeks	7 (15 items)	Sensation; mobility; emotion; cognition; self-care; pain; fertility	Varies by dimension	–0.03 – 1 (utility scores)	Canada	Yes
Health Utilities Index Questionnaire mark 3 (HUI-3) 15Q [49, 50]	Child self-report Adult self-report Proxy report	During past 4 weeks	8 (15 items)	Vision; hearing; speech; ambulation; dexterity; emotion; cognition; pain	Varies by dimension	–0.36 – 1 (utility scores)	Canada	Yes
Hospital Anxiety and Depression Scale (HADS) [36, 47]	Adult self-report	Last week	2 (14 items)	Anxiety; depression	0–3 rating scale	No total score calculated	UK	Yes
Individualized Neuromuscular Quality of Life Questionnaire (INQOL) [38]	Adult self-report	At the moment	10 + a total score (45 items)	Weakness; locking; pain; fatigue; activities;	7-point rating scale, varies by	Scoring unclear	UK	Yes

Table 1 Patient reported outcome measures assessing quality of life identified in published articles with samples of people with DMD (Continued)

PROM	Respondent type	Recall Period	N dimensions (items)	Dimensions of quality of life assessed	Response options	Total score range	Origin	Validated English copy and development papers freely available for review?
				independence; social relationships; emotions; body image; treatment; total	dimension			
KIDSCREEN-10 [51]	Child self-report Proxy report	Last week	1 total score (10 items)	Health-related quality of life	1–5 rating scale	10–50 (raw)	Multi-country	Yes
KIDSCREEN-27 [52]	Child self-report Proxy report	Last week	5 (27 items)	Physical well-being; psychological well-being; autonomy and parent relation; social support and peers; school environment	1–5 rating scale	No total score calculated	Multi-country	Yes
KIDSCREEN-52 [53]	Child self-report Proxy report	Last week	10 (52 items)	Physical well-being; psychological well-being; moods and emotions; self-perception; autonomy; parent relation and home life; financial resources; social support and peers; school environment; social acceptance (bullying)	1–5 rating scale	No total score calculated	Multi-country	Yes
Life Satisfaction Index for Adolescents (LSIA) [54–57]	Child and young adults self-report	At present	5 + a total score (45 items)	General well-being; interpersonal relationships; personal development; personal fulfilment; leisure and recreation; total	1–5 rating scale (plus 0 = N/A)	0–225 (raw)	Canada	Yes
Muscular Dystrophy Child Health Index of Life with Disabilities (MDCHILD) [8]	Child self-report	Past 4 weeks	7 + a total score (47 items)	Activities of daily living & independence; positioning, transferring, & mobility; comfort & endurance; emotions & behaviour; social interaction & school; health; your overall quality of life; total	Varies by dimension	0–100 (transformed)	Canada	Yes
Neurological Disorders Quality of Life Questionnaire (Neuro-QoL) [35]	Adult self-report	Varies by dimension	Up to 16 (up to 564 items in item banks)	Ability to participate in social roles and activities; anxiety; bowel function; cognitive function; communication; depression; emotional and behavioral dyscontrol; fatigue; lower extremity function – mobility; positive affect and well-being; satisfaction with social roles and activities;	1–5 rating scale	No total score calculated	USA	Yes

Table 1 Patient reported outcome measures assessing quality of life identified in published articles with samples of people with DMD (Continued)

PROM	Respondent type	Recall Period	N dimensions (items)	Dimensions of quality of life assessed	Response options	Total score range	Origin	Validated English copy and development papers freely available for review?
				sleep disturbance; sexual function; stigma; upper extremity function – fine motor, ADL; urinary/bladder function				
Offer Self-Image Questionnaire for Adolescents (OSIQ) [54, 57]	Child and young adult self-report or interview	Unknown/undefined	11 + a total score (130 items)	Impulse control; emotional tone; body and self-image; social relationships; morals; vocational and educational goals; family relationships; mastery of the external world; psychopathology; superior adjustment; total	1–6 rating scale	130–780 (raw)	USA	No
Patient Health Questionnaire 9-item (PHQ-9) [48]	Adult self-report	2 weeks	1 total score (9 items)	Depression	0–3 rating scale	0–27 (raw)	USA	Yes
Pediatric Neurological Disorders Quality of Life Questionnaire (Pediatric Neuro-QoL) [35]	Child self-report	Varies by dimension	Up to 11 (up to 161 items in item banks)	Anger; anxiety; cognitive function; depression; fatigue; lower extremity – mobility; pain; social relations – interaction with adults; social relations – interaction with peers; stigma; upper extremity – fine motor, ADL	1–5 rating scale	No total score calculated	USA	Yes
Pediatric Outcomes Data Collection Instrument (PODCI) [58–61]	Child self-report Proxy report	Varies by dimension	7 (86 items)	Global function & comfort; upper extremity function; physical function and sport; transfers and mobility; comfort; POSNA happy and satisfied; POSNA expectations	Varies by dimension	No total score calculated	USA	Yes
Pediatric Quality of Life Inventory (PedsQL) 3.0 DMD module [62, 63]	Child and young adult self-report Proxy report	Past month or past 7 days (acute version)	4 (18 items)	Daily activities; treatment barriers; worry; communication	0–4 rating scale	No total score calculated	USA	Yes
PedsQL 3.0 Multidimensional fatigue scale (MFS) [63, 64]	Adult self-report Child and young adult self-report Proxy report	Past month or past 7 days (acute version)	3 + a total score (18 items)	General fatigue; sleep/rest fatigue; cognitive fatigue; total fatigue	0–4 rating scale	0–100 (transformed)	USA	Yes
PedsQL 3.0 Neuromuscular module (NMM) [35, 50, 51, 63–69]	Child and young adult self-report	Past month or past 7 days (acute version)	3 + a total score (25 items)	About my/my child's neuromuscular disease; communication; about our	0–4 rating scale	0–100 (transformed)	USA	Yes

Table 1 Patient reported outcome measures assessing quality of life identified in published articles with samples of people with DMD (Continued)

PROM	Respondent type	Recall Period	N dimensions (items)	Dimensions of quality of life assessed	Response options	Total score range	Origin	Validated English copy and development papers freely available for review?
	Proxy report			family resources; total				
Pediatric Quality of Life Inventory (PedsQL) 4.0 Generic Core Scales (GCS) [35, 43, 51, 58, 61–65, 67, 70–77]	Adult self-report Child and young adult self-report Proxy report	Past month or past 7 days (acute version)	5 + a total score (23 items)	Physical health; psychosocial health; emotional functioning; social functioning; school functioning; total	0–4 rating scale	0–100 (transformed)	USA	Yes
PedsQL 4.0 Generic Short-form (SF-15) [78]	Adult self-report Child and young adult self-report Proxy report	Past month or past 7 days (acute version)	5 + a total score (15 items)	Physical health; psychosocial health; emotional functioning; social functioning; school functioning; total	0–4 rating scale	0–100 (transformed)	USA	Yes
Pittsburgh Sleep Quality Index (PSQI) [35, 74]	Adult self-report	Past month	8 + a total score (10 items)	Subjective sleep quality; sleep latency; sleep duration; sleep efficiency; sleep disturbance; use of sleep medication; daytime dysfunction; total	Varies by dimension	0–21 (raw)	USA	Yes
Satisfaction with Life Scale (SWLS) [48, 54]	Adult self-report	Undefined/present time	1 total score (5 items)	Life satisfaction	1–7 rating scale	5–35 (raw)	USA	Yes
Strength and Difficulties Questionnaire (SDQ) [79–81]	Child self-report Proxy report	Last 6 months	5 + a total score (25 items) + an impact supplement	Emotional symptoms; conduct problems; hyperactivity/inattention; peer relationship problems; total difficulties; prosocial behaviour	0–2 rating scale	0–40 (raw)	UK	Yes
Strips of Life with Emoticons Questionnaire (SOLE) [82]	Child self-report	Specific scenarios	1 total score (33 items)	Quality of life	0–2 rating scale (with pictures)	0–66 (raw)	Italy	No
TNO-AZL Children's Quality of Life questionnaire (TACQoL) [83]	Child self-report Proxy report	The last few weeks	7 (56 items)	Physical functioning; motor functioning; independent daily functioning; cognitive functioning and school performance; social contacts; positive moods; negative moods	Varies by dimension	No total score calculated	Netherlands	No
TNO-AZL Adult Quality of Life questionnaire (TAAQoL) [83]	Adult self-report	In the last month	12 (45 items)	Gross motor functioning; fine motor functioning; cognition; sleep; pain; social contacts; daily activities; sex; vitality; happiness; depressive mood; anger	Varies by dimension	No total score calculated	Netherlands	No

Table 1 Patient reported outcome measures assessing quality of life identified in published articles with samples of people with DMD (Continued)

PROM	Respondent type	Recall Period	N dimensions (items)	Dimensions of quality of life assessed	Response options	Total score range	Origin	Validated English copy and development papers freely available for review?
World Health Organisation Quality of Life Scale-Brief Version (WHOQOL-BREF) [33, 35–37]	Adult self-report or interview Proxy report	2 weeks	4 (26 items, 24 items make up domain scores)	Physical health; psychological; social relationships; environment	1–5 rating scale	No total score calculated	Multi-country	Yes

References next to PROM names represent published studies where the PROM has been used in a sample of people with DMD. PROM = patient reported outcome measure

not full-text primary research articles; and 1 did not meet the sample selection criteria. Of the remaining 23 records that had already met the selection criteria for Search A, 11 were excluded at full-text for containing information on measurement properties other than, and not including, content and structural validity; 4 were excluded as not containing information on measurement properties; and 3 for including a PROM that did not have a validated English copy that was free and/or available for review. The remaining 5 papers that met the selection criteria for Search B featured evidence on content validity ($n = 3$, of which one was classified as a development paper) and structural validity ($n = 2$). Finally, 33 PROM development papers were identified through a review of Google Scholar search results and 3 PROM development papers were identified through citation tracking, resulting in a final selection of 41 papers that met the selection criteria for Search B (see Fig. 1). These included 37 development papers, 2 content validity studies in DMD, and 2 structural validity studies in DMD.

The observed proportionate agreement between reviewers during selection, based on the primary database searches, was 92.4% at title/abstract, with Cohen's $\kappa = 0.51$ or "moderate agreement" and is similar to other published reviews [84, 85]. At full-text review, the observed proportionate agreement was 93.5% with Cohen's $\kappa = 0.82$ or "almost perfect agreement".

Following the searches, 26 PROMs were taken forward for COSMIN quality assessment on content and structural validity in DMD (Table 2). The remaining 14 PROMs were not assessed for the following reasons: a copy of the PROM itself and/or necessary development papers were not freely accessible for review (CAPE, CHQ-PF50, DISABKIDS Smileys, OSIQ, SF-36 v2); no formally validated English copy of the PROM was available or in use (AUQEL, DIKJ, DUC-25, SOLE, TAAQoL, TACQoL); the PROM was no longer available or recommended for use (BASC 1st edition, which has been superseded by the BASC 2); or it was unclear from the study which of a large number of possible variants of a PROM were used (pediatric Neuro-QoL, Neuro-QoL).

Content validity – appraisal of PROM development studies

Table 2 summarises key characteristics and COSMIN quality assessment of the development of the PROMs included in the review. Five PROMs were developed to be intended for use specifically within neuromuscular disorders (INQoL, PedsQL 3.0 NMM) or DMD (LSIA, MDCHILD, PedsQL 3.0 DMD module). Eleven PROMs either had no patients involved in their development, or it was unclear if patients were involved.

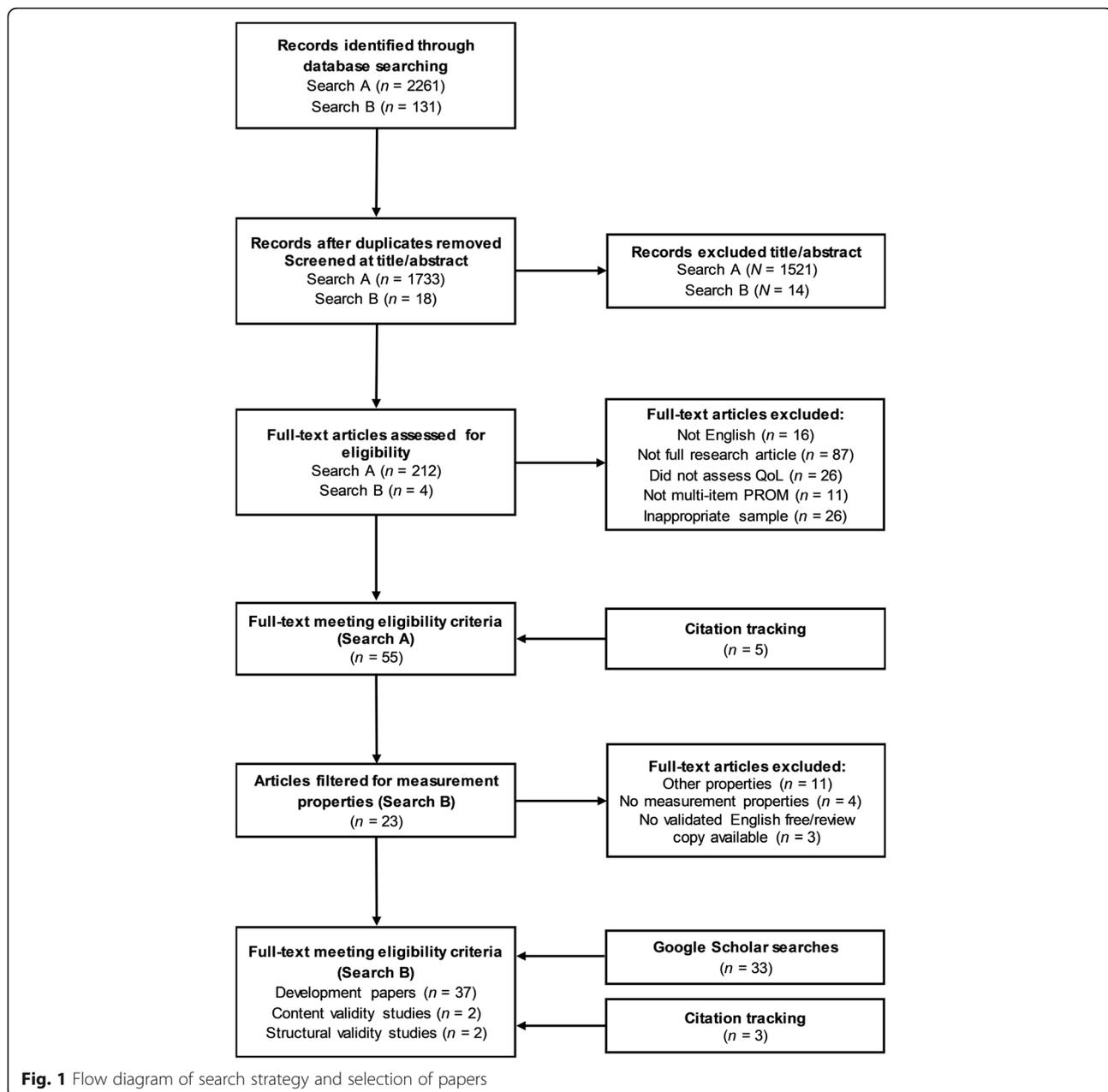
The joint most common COSMIN quality rating assigned to the PROMs for concept elicitation was inadequate ($n = 12$). This was primarily due to: the PROM development study not being performed in a sample of patients representing the target population (BDI, EQ-5D-3L, GAD-7, HADS, HUI 15Q, PedsQL 3.0 MFS, PHQ-9, SDQ, SF-36, and SWLS); or inadequacies within the details of the qualitative methods used (FSS, INQoL). The concept elicitation study of 11 further PROMs was rated as doubtful due to at least some unclear details/suspected problems within the qualitative methods used (CALI, DCGM-37, LSIA, MDCHILD, PODCI, PedsQL 3.0 NMM, PedsQL 3.0 DMD, PedsQL 4.0 GCS, PedsQL 4.0 SF-15, PSQI, WHOQOL-BREF). Only the KIDS CREEN family of measures ($n = 3$) received an adequate rating for concept elicitation and PROM design. However, the KIDSCREEN measures received a doubtful rating for the overall PROM development study, for failing to provide evidence that comprehensibility and comprehensiveness were assessed in the cognitive interview/pilot study of the PROM.

Content validity – appraisal of content validity studies

Only 2 published articles had independently assessed the content validity of the QoL PROMs in samples of people with DMD (Table 3). Neither of these studies were conducted in an English language context, and instead were cross-cultural validation studies. Hu et al. (2013) [67] assessed the relevance, comprehensiveness, and comprehensibility of the PedsQL 3.0 NMM in Chinese children with DMD. Simon et al. (2017) [56] assessed comprehensibility of the LSIA in Brazilian children with DMD, and comprehensiveness in professionals. However, both of these studies received ratings of doubtful due to at least some unclear details/suspected problems within the qualitative methods used.

Content validity evidence synthesis

The evidence from the PROM development papers and content validity studies was combined with reviewer ratings of the PROMs to produce a synthesis of the available evidence using the 10 COSMIN criteria for good content validity [13]. Most of the quality of the evidence was downgraded from High to Low or Very Low due to the assessment being based on development studies of doubtful or inadequate quality, respectively [13]. Only the LSIA and the PedsQL 3.0 NMM had moderate supporting evidence, featuring independent content validity studies as well as development papers. The KIDSCREEN measures and the LSIA were the only PROMs to receive satisfactory results for all three dimensions of content validity: relevance; comprehensiveness; and comprehensibility, based on the evidence available. Full synthesised results are presented in Table 4.



Structural validity - appraisal of structural validity studies

Two studies had assessed the structural validity of the PROMs included in this review in samples of people with DMD (Table 5). Both of these were conducted using English versions of the PROMs and either in the UK or USA. Lim et al. (2014) [72] assessed the structural validity of the PedsQL 4.0 GCS using an unspecified Rasch model in 63 boys with DMD. This study received a COSMIN quality rating of doubtful because it was doubtful that the sample size included in the analysis was adequate. Landfeldt et al. (2018) [66] assessed the structural validity of the PedsQL 3.0 NMM using a Rasch partial-credit model (PCM) in 278 people with

DMD. This study received a very good COSMIN quality rating for its methodological content.

Structural validity evidence synthesis

Of the 2 studies that assessed the structural validity of the PedsQL 4.0 GCS and PedsQL 3.0 NMM in people with DMD, neither provided satisfactory results (Table 4). First, the structural validity of the PedsQL 4.0 GCS in people with DMD received an indeterminate rating, as key details of the results from the Rasch model denoting good measurement properties were not reported. Due to the risk of bias assessment of Lim et al. (2014) [72] the quality of the evidence supporting this indeterminate

Table 2 Characteristics and assessment of development papers for measures included in the review

PROM	Reference(s)	Original language	Construct definition	Target population	Intended context of use	Concept elicitation study	
						COSMIN quality rating	Were patients involved?
BDI	Beck et al. 1961 [86]	English (US)	"the items were chosen on the basis of their relationship to the overt behavioral manifestations of depression and do not reflect any theory regarding the etiology or the underlying psychological processes in depression"	Adult patients with suspected symptoms of depression	Quantitative assessment of the intensity of depression in diagnostic and research settings	Inadequate	No
CALI	Palermo et al. 2004 [87]	English (US)	"functional impairment, defined as difficulty in performing age-appropriate physical, mental, and social activities in daily life due to physical health status (...) functional impairment due to pain (...) specific areas of functioning that are important to children and adolescents with recurrent and chronic pain"	School-age children and adolescents with recurrent and chronic pain	Research and clinical care	Doubtful	Yes
DCGM-37	Petersen et al. 2005 [88] Ravens-Sieberer et al. 2007 [89]	English (UK)	"a multidimensional construct with social, physical, emotional, and functional domains"	Children aged 4–7 years and 8–16 years with chronic health conditions	Clinical studies or surveys	Doubtful	Yes
EQ-5D-3L ^a	EuroQol Group 1990 [6] Brooks et al. 1996 [7]	Multiple, including English (UK)	"Health-related quality of life"	"Large-scale surveys of the community and (...) for use in postal surveys"	"Complement other quality of life measures, collection of common data set for reference. Generate cross-national comparisons of health state valuations."	Inadequate	No
FSS	Krupp et al. 1989 [90]	English (US)	"Fatigue"	Patients with "clinical disorders"	Clinical research studies and surveys	Inadequate	No
GAD-7	Spitzer et al. 2006 [91]	English (US)	"We first selected potential items for a brief GAD [Generalized Anxiety Disorder] scale (...) that reflected all of the <i>Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)</i> symptom criteria for GAD and (...) on the basis of review of existing anxiety scales."	General adult population	Clinical practice and research	Inadequate	No
HADS	Zigmond & Snaith 1983 [92]	English (UK)	"depression subscale were largely based on the anhedonic state (...) psychic manifestations of anxiety neurosis"	Patients under investigation and treatment in medical and surgical departments in non-psychiatric hospital departments	Clinical/screening use within non-psychiatric hospital departments	Inadequate	No
HUI-2 / HUI-3 (15Q)	Feeny et al. 1995 [93] Torrance et al. 1996 [94]	English (US)	"The HUI Mark II and Mark III systems are based on concepts of functional capacity rather than performance (...) generic health profile measures that	Originally survivors of childhood cancer (HUI-2), extended to adults	Clinical evaluative and population health survey studies, in clinical trials, and cost-utility analyses	Inadequate	Unknown

Table 2 Characteristics and assessment of development papers for measures included in the review (Continued)

PROM	Reference(s)	Original language	Construct definition	Target population	Intended context of use	Concept elicitation study	
						COSMIN quality rating	Were patients involved?
			also permit the computation of a single summary score quantifying health-related quality of life"				
INQoL	Vincent et al. 2007 [95]	English (UK)	"The structure of INQoL was based on the ICIDH-2 model of disease incorporating the concepts of Impairment, Activities, and Participation."	Adults with neuromuscular disorders (16+ years)	Clinical and research use	Inadequate	Yes
KIDSCREE N-52	Ravens-Sieberer et al. 2001 [96] Ravens-Sieberer et al. 2005 [97] Detmar et al. 2006 [98]	Multiple, including English (UK)	"Health-related quality of life is described as a multidimensional construct covering physical, emotional, mental, social, and behavioral components of well-being and function as perceived by patients and/or individuals (...) agreement was reached that the questionnaire should aim to measure HRQOL as a generic construct in largely healthy children, thus more emphasis was given to the inclusion of psychosocial domains, and less to domains of physical functioning or symptoms such as pain."	Healthy and chronically-ill children and adolescents between 8 and 18 years	Epidemiological and paediatric studies, clinical settings (healthcare system), and health services research	Adequate	Yes
KIDSCREE N-27	Ravens-Sieberer et al. 2006 [99]	Assumed the same as KIDS CREEN-52	Assumed the same as KIDS CREEN-52	Assumed the same as KIDS CREEN-52	Assumed the same as KIDSCREEN-52	Adequate	Yes
KIDSCREE N-10	Ravens-Sieberer et al. 2006 [99]	Assumed the same as KIDS CREEN-52	Assumed the same as KIDS CREEN-52	Assumed the same as KIDS CREEN-52	Assumed the same as KIDSCREEN-52	Adequate	Yes
LSIA	Reid & Renwick 1994 [54]	English (US)	"quality of life is to conceptualize it as a subjective phenomenon. Specifically, it is viewed in terms of the individual's feelings and evaluations of his or her life circumstances. Many researchers who study quality of life within this perspective emphasize the importance of measuring the individual's degree of life satisfaction. In other words, they are interested in how pleased an individual feels about particular aspects of his or her life"	"Individuals between the ages of 12 and 19 years who have DMD"	Research instrument and potentially useful as a clinical measure	Doubtful	Yes
MDCHILD	Propp, 2017 [100] Propp et al.	English (UK)	"Health-related priorities for children with DMD (...) defined as concerns,	Children with DMD (assumed 5–18 years)	Cohort studies, clinical trials, and clinical decision-making	Doubtful	Yes

Table 2 Characteristics and assessment of development papers for measures included in the review (Continued)

PROM	Reference(s)	Original language	Construct definition	Target population	Intended context of use	Concept elicitation study	
						COSMIN quality rating	Were patients involved?
	2019 [8]		desires, and expectations arising from the lived experience of that condition"				
PedsQL 3.0 DMD	Uzark et al. 2012 [62]	English (US)	"Health-related quality of life (QoL), a multidimensional construct that includes physical, psychological, and social functioning, has emerged as an important outcome in pediatric populations with chronic health conditions."	Children with DMD from 2 to 18 years	Assumed the same as PedsQL 4.0 GCS	Doubtful	Yes
PedsQL 3.0 MFS	Varni et al. 2002 [101]	English (US)	"designed to measure child and parent perceptions of fatigue in pediatric patients"	Assumed the same as PedsQL 4.0 GCS	"may be utilized as outcome measures in pediatric cancer clinical trials, research, and clinical practice for HRQOL"	Inadequate	Yes
PedsQL 3.0 NMM	Iannaccone et al. 2009 [102]	English (US)	"HRQOL is a multidimensional construct, consisting at the minimum of physical, psychological (including emotional and cognitive), and social health dimensions delineated by the World Health Organization. HRQOL has emerged as the most appropriate term for quality of life dimensions that represent a patient's perceptions of the impact of an illness and its treatment on their own functioning and well-being and which are within the scope of healthcare services and medical products."	Children and young people with neuromuscular disorders, in particular spinal muscular atrophy	Assumed the same as PedsQL 4.0 GCS	Doubtful	Yes
PedsQL 4.0 GCS	Varni et al. 1999 [103]	English (US)	"The PedsQL measures the patient's and the parent's perceptions of the patient's HRQOL, as defined in terms of the impact of disease and treatment on an individual's physical, psychological, and social functioning, and by disease/treatment-specific symptoms."	Children aged 8–18 across various pediatric chronic health conditions	Epidemiological studies, clinical trials, and performance improvement studies	Doubtful	Yes
PedsQL 4.0 SF-15	Varni et al. 1999 [103]	English (US)	Assumed the same as PedsQL 4.0 GCS	Assumed the same as PedsQL 4.0 GCS	Assumed the same as PedsQL 4.0 GCS	Doubtful	Yes
PHQ-9	Spitzer et al. 1999 [104] Kroenke et al. 2001 [105]	English (US)	"Depression (...) using diagnostic criteria from the <i>Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition (DSM-III-R)</i> and <i>Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)</i> ."	General adult population	Clinical practice and research	Inadequate	No

Table 2 Characteristics and assessment of development papers for measures included in the review (Continued)

PROM	Reference(s)	Original language	Construct definition	Target population	Intended context of use	Concept elicitation study	
						COSMIN quality rating	Were patients involved?
PODCI	Daltroy et al. 1998 [106]	English (US)	"The POSNA outcomes instrument scales assess upper extremity function, transfers and mobility, physical function and sports, comfort (painfree), happiness and satisfaction, and expectations for treatment. A POSNA global scale combines the three function subscales and comfort."	Children aged 2–18 years with musculoskeletal disorders	"Patient-based instrument"	Doubtful	Yes (assumed)
PSQI	Buysse et al. 1989 [107]	English (US)	"sleep quality is a readily accepted clinical construct, it represents a complex phenomenon that is difficult to define and measure objectively. 'Sleep quality' includes quantitative aspects of sleep, such as sleep duration, sleep latency, or number of arousals, as well as more purely subjective aspects, such as "depth" or "restfulness" of sleep"	Clinical/psychiatric populations	Psychiatric clinical practice and research activities	Doubtful	No
SDQ	Goodman 1997 [108]	English (UK)	"young people's behaviours, emotions, and relationships"	Children and young people (aged 4–16 years)	"to meet the needs of researchers, clinicians, and educationalists"	Inadequate	No
SF-36 v1.0 ^a	Ware & Sherbourne 1992 [109] Hays et al. 1993 [110] Jenkinson et al. 1999 [111] Ware 2000 [112]	English (US)	"'Health', eight concepts: physical functioning, social and role functioning, mental health, general health perceptions, bodily pain, and vitality."	"General population and patients"	"Clinical practice and research, healthy policy evaluations, and general population surveys"	Inadequate	No
SWLS	Diener et al. 1985 [113]	English (US)	"Life satisfaction refers to a cognitive, judgmental process. Shin and Johnson (1978) define life satisfaction as "a global assessment of a person's quality of life according to his chosen criteria" (p. 478)"	Unclear	Unclear	Inadequate	No
WHOQOL-BREF	WHOQOL Group 1994 [114] WHOQOL Group 1995 [115] Skevington et al. 1997 [116] WHOQOL Group 1998 [117]	Multiple, including English (UK)	"It is a broad ranging concept incorporating, in a complex way, the person's physical health, psychological state, level of independence, social relationships, personal beliefs, and relationship to salient features of the environment (...) At minimum, quality of life includes the following dimensions: physical (individuals' perception of	"assess the quality of life of chronic disease sufferers, informal caregivers of the sick and disabled, people living in high-stress conditions like refugees, and 'healthy' people"	"in routine clinical work, large scale epidemiological studies and in clinical trials"	Doubtful	Yes

Table 2 Characteristics and assessment of development papers for measures included in the review (Continued)

PROM	Reference(s)	Original language	Construct definition	Target population	Intended context of use	Concept elicitation study	
						COSMIN quality rating	Were patients involved?
			their physical state), psychological (individuals' perception of their cognitive and affective state) and social (individuals' perception of the interpersonal relationship relationships and social roles in their life). (...) The WHOQOL includes a spiritual dimension (the person's perception of 'meaning in life', or the overarching personal beliefs that structure and qualify experience)."				

^aPROM development information from prior COSMIN review [118], not re-extracted or re-rated in this review, based on COSMIN guidance [13]. PROM = patient reported outcome measure; COSMIN = COnsensus-based Standards for the selection of health Measurement INstruments

conclusion was rated as very low. Second, the structural validity of the PedsQL 3.0 NMM in people with DMD received an unsatisfactory rating, as the psychometric criteria for good measurement properties were not met. The favourable risk of bias assessment for Landfeldt et al. (2018) [66] meant that the quality of evidence supporting this conclusion was graded as high.

Quality assurance of the review

The quality of this review was self-assessed against a newly derived COSMIN checklist [22], designed to evaluate the quality of systematic reviews of health-related PROMs. The results are displayed in Additional file 3.

In general, the review meets numerous quality indicators as defined by the COSMIN team, including the elements included in the research aim, search strategies, article selection, and assessment of measurement properties and quality. In a couple of instances, criteria have been partly met. For example, in this review all instruments were included where a *validated English copy was freely available for review*. It is possible that additional instruments could have been included if licenses were paid for to access the relevant PROMs and development materials, and this could be considered a limitation. Second, citation tracking (i.e. reference checking) was conducted on the final set of articles eligible at Stage 2

Table 3 Characteristics and assessment of content validity papers in DMD samples for measures included in the review

PROM	Reference	Language (Country)	DMD sample characteristics			COSMIN rating			Results (summary)
			N	Age (years, \pm SD)	% ambulatory	Relevance	Comprehensiveness	Comprehensibility	
LSIA	Simon et al. 2017 [56]	Brazilian Portuguese (Brazil)	43	11.4 \pm 3.38	Not stated	/	Doubtful	Doubtful	The level of comprehension reached via the final Probe technique was 97% for the parent version and 95% for the patient version, which is above the minimum of 85% required.
PedsQL 3.0 NMM	Hu et al. 2013 [67]	Chinese (China)	56	7.54 \pm 4.06	37 children "could climb stairs"	Doubtful	Doubtful	Doubtful	Cognitive debriefing was conducted with six children with DMD and their parents to confirm that the final Chinese version was understandable and acceptable.

/ = content validity aspect not evaluated. PROM = patient reported outcome measure; COSMIN = COnsensus-based Standards for the selection of health Measurement INstruments

Table 4 Evidence synthesis on the content and structural validity of measures that have been used to assess quality of life in people with DMD

PROM	Content validity			Quality of evidence	Structural Validity	
	Relevance	Comprehensiveness	Comprehensibility		Rating of results	Quality of evidence
BDI	±	–	±	Very low	?	?
CALI	+	–	±	Low	?	?
DCGM-37	±	+	+	Low	?	?
EQ-5D-3L	+	–	+	Very low	?	?
FSS	±	–	±	Very low	?	?
GAD-7	+	–	+	Very low	?	?
HADS	–	–	–	Very low	?	?
HUI-2 / HUI-3 (15Q)	–	–	–	Very low	?	?
INQoL	±	±	+	Very low	?	?
KIDSCREEN-52	+	+	+	Low	?	?
KIDSCREEN-27	+	+	+	Low	?	?
KIDSCREEN-10	+	+	+	Low	?	?
LSIA	+	+	+	Moderate	?	?
MDCHILD	±	+	+	Low	?	?
PedsQL 3.0 DMD	±	?	±	Very low	?	?
PedsQL 3.0 MFS	±	–	±	Very low	?	?
PedsQL 3.0 NMM	±	?	±	Moderate	–	High
PedsQL 4.0 GCS	±	+	±	Low	?	Very low
PedsQL 4.0 SF-15	±	+	±	Low	?	?
PHQ-9	+	–	±	Very low	?	?
PODCI	±	+	±	Very low	?	?
PSQI	±	–	±	Very low	?	?
SDQ	–	–	+	Very low	?	?
SF-36 v1.0	±	+	±	Very low	?	?
SWLS	–	–	±	Very low	?	?
WHOQOL-BREF	+	+	±	Very low	?	?

+ = satisfactory results; – = unsatisfactory results; ± = inconsistent results; ? = indeterminate. PROM = patient reported outcome measure

of the searches ($n = 41$), but not on results eligible for inclusion at Stage 1.

Discussion

In this systematic review, the published scientific evidence on the content and structural validity of PROMs used to measure at least one aspect of QoL in people with DMD was thoroughly evaluated. The overriding theme was one of sparse evidence. Many PROMs that are being used to assess aspects of QoL in people with DMD are being utilised without the accompanying good quality evidence that supports their validity for this task. Only five of the PROMs uncovered in this review were specifically designed for use in people with neuromuscular problems (three for DMD), and only two of these have had their content and/or structural validity independently assessed in this population (with the content

validity studies involving translated versions). When the evidence is available, most of it is either of a low quality, featuring insufficient detail in the published articles to make thorough and comprehensive assessments of content and structural validity as demanded by COSMIN [16], leading to doubtful ratings. Indeed, one of the highest quality pieces of evidence reviewed in terms of reported methodology, Landfeldt et al. (2018) [66], reported insufficient structural validity of the PedsQL 3.0 Neuromuscular module (NMM) in DMD.

The results from the review may not be viewed as surprising. Many of the PROMs identified are what could be described as “legacy” measures. They were developed at a time when the science of construct and item generation was largely overlooked. The content of instruments was largely defined by clinical or expert opinion, with little explanation of what that entailed. The reporting of

Table 5 Characteristics, assessment, and results of structural validity papers in DMD samples for measures included in the review

PROM	Reference	Country (language)	Patient characteristics				COSMIN Quality Rating	Analysis – model	Results (summary)
			N	Age (yr, M ± SD)	% ambulatory	PROM score			
PedsQL 4.0 GCS	Lim et al. 2014 [72]	USA (English)	63 boys with DMD (and up to 50 parents, not necessarily matched)	10.2 ± 2.5	95.24	Child: M = 64.5, SD = 15.3. Parent: M = 56.2, SD = 12.9.	Doubtful	Rasch (model not specified)	Model misfit for items determined with infit > 1.4 and outfit > 2.0 MnSq values and standardized scores > 2.0. All items fit in parent proxy-reports of physical health scale and child self-reports of psychosocial health scale. 2 out of 8 items showed high infit statistics in child self-reports of the physical health scale (taking a bath or shower; doing chores around the house). In addition 2 out of 15 items showed high infit for the parent proxy-reports of the psychosocial health scale (trouble sleeping; keep up with school work).
PedsQL 3.0 NMM	Landfeldt et al. 2018 [66]	UK / USA (English)	278 (95 UK)	16 ± 7	40% not “full-time wheelchair dependent”	Not reported	Very good	Rasch PCM	Eight items displayed inadequate fit ($\chi^2: p > 0.01$). Six items had fit residuals ≤ -2.5 or ≥ 2.5 (4 significant at $p < .05$). Inadequate overall fit (χ^2 item-trait interaction: $p < .001$). Disordered thresholds for 22 of 25 items. Suboptimal targeting.

PROM = patient reported outcome measure; COSMIN = COnsensus-based Standards for the selection of health Measurement INstruments

such stages in publications or questionnaire manuals was not commonplace. The transparency of reporting on the early stages of PROM development has only gained traction in the last decade or so. Whilst this is a positive step for researchers, clinicians and users alike, progress can be limited by journal restrictions on word count and remit. It is however possible for such legacy measures to be appropriately validated (or have their validity assessed) in properly designed studies assessing content or structural validity in modern samples of people with DMD. The problem observed in this review is that researchers are likely using such measures as a consequence of precedent or tradition, rather than a supportive evidence base.

Another related legacy issue within PROM development, which this review touches upon and has changed for the better over time, is a recognition of the importance of direct patient involvement in developing PROMs [119, 120]. In this review, almost half (11 out of 26) of the PROMs did not demonstrate any evidence of patient involvement in their development. While most of these PROMs are legacy measures, this is a noteworthy figure, given that patient involvement is the only way to ensure a PROM is capturing health and QoL outcomes in a way that is relevant, comprehensive, and comprehensible to the patient population [119]. The use of patient involvement in PROM development is thus advantageous for researchers and patients alike. To help guide PROM developers, a recent framework has been published to help researchers fully incorporate patient and public involvement (PPI) in the development of PROMs moving forward [120].

In the current review, some PROMs performed better than others under COSMIN assessment. First, the KIDS CREEN instrument (all versions) does show some evidence of applicability given that it covers many aspects of QoL. The PROM development study for the KIDS CREEN instrument was the only one rated as adequate, it was designed to assess QoL in children and adolescents with chronic illnesses, and the ratings for the content validity of the measure were positive (based on the available evidence in the measure's development). However, it must also be borne in mind that there is little or no direct evidence to support the content or structural validity in DMD, specifically. The original KIDSCREEN instrument (52-item version) was designed to assess multiple aspects of QoL, namely: physical well-being; psychological well-being; moods and emotions; self-perception; autonomy; parent relation and home life; financial resources; social support and peers; school environment; and social acceptance (bullying), covering much of the CMQM framework [9]. The conceptual framework of the instrument is thus intuitively applicable to the Duchenne community; however the

measurement of impact may be limited due to the target age range of the PROM itself (8–18 years). While this is not uncommon (i.e. differences in measuring QoL from child to adulthood), there is some question of the applicability for the broader DMD population given the lower age target.

The second-best performing PROM in this review was the LSIA, which received a satisfactory score for relevance, comprehensiveness, and comprehensibility in terms of content validity, based on the information available and reviewers' ratings of the PROM itself. However, the development study for this paper lacked key details necessary in good PROM development, and thus was rated as doubtful. Furthermore, while the LSIA was one of few measures to feature a content validity study, it was a cross-cultural adaptation study of a Brazilian version of the measure, and the results of the formal assessment of this study were doubtful. While the measure is comprehensive, it only comes in a 45-item version, which is potentially quite burdensome. Furthermore, the measure is designed for use in children and young adults only, and may not generalise to adults with DMD.

The most recent PROM developed specifically for use in children and adolescents with DMD was the MDCH ILD. Although the PROM is designed to measure "health-related priorities" [8], much of the content maps onto the CMQM framework [9] and thus covers QoL. While the MDCHILD had many commendable strengths in PROM design, the overall rating of the PROM development, based on the COSMIN worst score counts system [25], was rated as doubtful due to lack of details reported in the development papers. For example, it was unclear if skilled interviewer(s) were used; to what degree data was coded independently; and to what degree, if at all, at least two researchers were involved in the data analysis. This led to a low quality of evidence. Further, because the target population of interest was not clearly defined (i.e. age ranges were not specified), despite performing well in other areas, the PROM received an inconsistent rating for relevance. These results speak to the potential harshness of a worst score counts system advocated by COSMIN, which we discuss further below. Further, because the PROM is new, there is a lack of published content validity studies that may improve the quality of evidence for the MDCHILD going forward, such as that contained in a non-peer-reviewed thesis [100], not eligible for inclusion in the current review.

The PedsQL and associated modules were the most commonly used out of all the PROMs identified within the review. It should be noted that the development studies of the PedsQL were rated as doubtful. There was little evidence to support the content validity of the neuromuscular module of the PedsQL 3.0 (NMM). Furthermore, the psychometric properties of the NMM

were not well supported by Landfeldt et al. (2018) [66]. The inclusion of PedsQL within clinical practice, cohort studies or pragmatic trials in DMD thus appears to be based upon precedent and common use, rather than published empirical evidence of suitability, based on content and structural validity. A notable advantage of the PedsQL (and its derivatives) is the young child (via proxy report), child (self and proxy report), young adult forms (self-report), and adult forms, which have now been developed. A further consideration is that the PedsQL scales are designed to be used in parallel (e.g. the generic core scales with the NMM or DMD modules), but were assessed individually under COSMIN guidance. Thus comprehensiveness may be improved by using these scales together.

The search identified some PROM instruments that we were unable to obtain. Access to the PROM and/or associated development papers was limited due to licensing requirements, and therefore it was not possible to include these instruments within the review. It is unlikely that these instruments are commonly used within research and/or clinical practice due to the difficulties around access. Their suitability for the DMD population cannot formally be determined; however, their use is likely to be limited by a lack of accessibility derived from license restrictions, reflected in the few citations in which they appeared.

This review adopted guidance developed by the COSMIN initiative, and has adhered to their recommended methods in identification of evidence, data extraction, data assessment and data synthesis. Whilst the appropriateness of these robust methods cannot be questioned, this has resulted in relatively low ratings of the PROMs included within the review. It is important to recognise that this does not suggest categorically that the instruments used within published and/or current studies are not appropriate or fit for purpose; content and structural validity only form one component of PROM suitability within a population. Furthermore, as stated, many of the instruments were developed at a time when instrument development methods and procedures were not reported – that is not to say the development of the instruments is flawed, just that an assessment of them cannot be made. The COSMIN appraisal tools assume a worst score counts system for the rating of the methodological quality of studies [25]. This means that, in theory, a study could be rated as very good or adequate on all but one criteria, on which it is rated as doubtful or inadequate, and the overall score is thus reduced to the latter lower-quality rating. Sometimes this can be because key details, such as whether skilled interviewers were used, are not reported.

This review is not without its limitations. While the methodological approach of the review is robust and

follows the recommendations of COSMIN and that of other published reviews, it must be acknowledged that the rating criteria of the PROMs identified can be viewed as harsh. The COSMIN approach encourages researchers and reviewers to critically appraise evidence of PROM development – however the presence of evidence within published literature is sparse. That is not to say that the development phases did not occur, merely that they are not reported and/or not reported in sufficient detail as required by COSMIN assessment. To critique a PROM's applicability using this criterion could be perceived as being unduly critical; more recent PROMs tend to report the early stages of instrument development, and we are assessing all PROMs by modern standards. Similarly, the descriptions of PROMs themselves are often lacking. Basic information such as number of items, recall period, domain structure and scoring procedure were noted to be sporadically reported, although better in recent literature. The COSMIN-recommended reviewer rating of the identified PROMs for suitability for DMD (as reported in Table 4) has a large subjective component. Whilst this was completed as per the COSMIN guidelines (with two reviewers and discrepancies reconciled following discussion), some of the ratings are at risk of bias based on the team of raters (i.e. QoL researchers). For example, it is not known whether similar ratings of suitability would be achieved if reviewed by an individual with DMD, a family member or carer of a person with DMD, or a clinician, and we recommend that PPI is incorporated in future COSMIN reviews of content validity. This is further exacerbated when we consider what QoL is – for the purpose of this review it was a multidimensional construct, PROMs that measure a subset of interest (such as depression) may be appropriate to include within studies as part of a host/suite of measures.

The focus of this review was to report on the content and structural validity of PROM instruments that have been used to quantify the impact of DMD on individuals' QoL. However, content and structural validity only address some aspects of PROM suitability, and further work could be undertaken to formally appraise the instruments described. Other measurement properties, such as psychometric performance, could be considered. Given that DMD is a rare condition, the development and validation of PROMs that measure the impact of the condition on QoL is challenging. The number of participants included within various phases of PROM development and validation will be lower than that of a condition such as diabetes, asthma or eczema. Accordingly, the inclusion of subsidiary samples such as other neuromuscular disorders, may be of interest. However, it is not known how appropriate this would be. It can be postulated that other neuromuscular disorders could

imply similar impacts upon QoL, however this has not been explored within the context of this review.

Conclusions

In conclusion, evidence on the content and structural validity of PROMs assessing QoL in DMD is lacking. Accordingly, our first recommendation from this review is for more research into the content and structural validity of QoL PROMs used in DMD, and, if PROMs are found to be insufficient on these criteria, for additional PROM development within DMD. Second, as the result of this COSMIN assessment, without further direct content validation work in DMD, we would provisionally recommend the KIDSCREEN for measuring QoL in children and adolescents with DMD. Nonetheless, we caution that the KIDSCREEN has not been formally validated in samples of people with DMD. Accordingly, more research is needed to definitively support the continued use of KIDSCREEN (and its derivatives) within DMD. Finally, in the absence of further evidence, it is difficult to recommend the routine use of a measure to assess QoL in adults with DMD on content and structural grounds. Instead, the findings of this review support the need for further PROM development, which is able to accurately assess the impact of DMD on QoL.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12955-020-01511-z>.

Additional file 1. Full search strategies. Tables showing the full search strategy used in the review.

Additional file 2. Example COSMIN ratings KIDSCREEN-52. Example COSMIN content validity rating spreadsheet for the KIDSCREEN-52.

Additional file 3. Quality assurance of the review. Quality assessment of the systematic review against COSMIN guidance.

Abbreviations

AUQEL: Autoquestionnaire Qualité de vie Enfant Imagé; BASC: Behavior Assessment System for Children; BDI: Beck Depression Inventory; CALI: Child Activity Limitations Interview; CAPE: Children's Assessment of Participation and Enjoyment; CFA: Confirmatory factor analysis; CFI: Comparative fit index; CHQ-PF50: Child Health Questionnaire - Parent Form 50; CMQM: Comprehensive Model of QoL in Muscular Dystrophy; COSMIN: COnsensus- based Standards for the selection of health Measurement INstruments; DCGM-37: DISABKIDS generic module; DIKJ: Depressions-Inventar für Kinder und Jugendliche; DMD: Duchenne muscular dystrophy; DUC-25: Dutch Children AZL/TNO Questionnaire Quality of Life Short Form; EQ-5D(- 3 L): Euroqol instrument, Euroqol-5 dimension (3 level); EQ-5D-Y: Euroqol instrument, Euroqol-5 dimension (youth version); FSS: Fatigue Severity Scale; GAD-7: Generalized Anxiety Disorder Scale 7-item; GRADE: Grading of Recommendations, Assessment, Development and Evaluations; HADS: Hospital Anxiety and Depression Scale; HUI: Health Utilities Index; INQOL: Individualized Neuromuscular Quality of Life Questionnaire; IRT: Item response theory; LSIA: Life Satisfaction Index for Adolescents; MDCHILD: Muscular Dystrophy Child Health Index of Life with Disabilities; Neuro-QoL: Neurological Disorders Quality of Life Questionnaire; OSIQ: Offer Self-Image Questionnaire; PBM: Preference based measure; PedsQL 3.0 DMD: Pediatric Quality of Life Inventory (PedsQL) 3.0 DMD module; PedsQL 3.0 MFS: PedsQL 3.0 Multidimensional fatigue scale; PedsQL 3.0 NMM: PedsQL 3.0 Neuromuscular module; PedsQL 4.0 GCS: Pediatric

Quality of Life Inventory (PedsQL) 4.0 Generic Core Scales; PedsQL 4.0 SF-15: PedsQL 4.0 Generic Short-form; PHQ-9: Patient Health Questionnaire Depression Scale; PODCI: Pediatric Outcomes Data Collection Instrument; PROM: Patient reported outcome measure; PROSPERO: International Prospective Register of Systematic Reviews; PSQI: Pittsburgh Sleep Quality Index; QoL: Quality of life; Rasch PCM: Rasch Partial-credit model; RMSEA: Root mean square error of approximation; SDQ: Strengths and Difficulties Questionnaire; SF-36: 36-item Short-Form Health Survey; SOLE: Strips of Life with Emoticons Questionnaire; SRMR: Standardized root mean squared residual; SWLS: Satisfaction with Life Scale; TAAQoL: TNO-AZL Questionnaire for Adult's Health-Related Quality of Life; TACQOL: TNO-AZL Questionnaire for Children's Health-Related Quality of Life; TLI: Tucker Lewis index; WHOQOL-BREF: World Health Organisation Quality of Life Scale-Brief Version

Acknowledgements

Not applicable.

Authors' contributions

PP and JC contributed to the conception and design of the work. HBW contributed to the systematic search strategy and conducted the searches. PP, JC, and PM contributed to data extraction, reviewing, and rating. PP drafted the manuscript. PP and JC substantively revised the manuscript. All authors read and approved the submitted version of the manuscript.

Funding

This project was funded by project HERCULES. 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. The funders provided comments on an earlier draft of the manuscript, but had no further role in study design, analysis, or interpretation.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Two of the authors (PP and JC) are employed on a related project funded by Duchenne UK that involves developing a new preference-based patient reported outcome measure of quality of life for people with DMD.

Received: 10 July 2019 Accepted: 24 July 2020

Published online: 03 August 2020

References

- Mendell JR, Shilling C, Leslie ND, Flanigan KM, al-Dahhak R, Gastier-Foster J, et al. Evidence-based path to newborn screening for Duchenne muscular dystrophy. *Ann Neurol*. 2012;71(3):304–13.
- Bradley DM, Parsons EP, Clarke AJ. Experience with screening newborns for Duchenne muscular dystrophy in Wales. *Bmj*. 1993;306(6874):357–60.
- Strehle E-M, Straub V. Recent advances in the management of Duchenne muscular dystrophy. *Arch Dis Child*. 2015;100(12):1173–7.
- Kieny P, Chollet S, Delalande P, Le Fort M, Magot A, Peroon Y, et al. Evolution of life expectancy of patients with Duchenne muscular dystrophy at AFM Yolaine de Kepper Centre between 1981 and 2011. *Ann Phys Rehabil Med*. 2013;56(6):443–54.
- Landfeldt E, Lindgren P, Bell CF, Schmitt C, Guglieri M, Straub V, et al. Compliance to care guidelines for Duchenne muscular dystrophy. *J Neuromuscul Dis*. 2015;2(1):63–72.
- EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. *Health policy*. 1990;16(3):199–208.
- Brooks R, EuroQol group. EuroQol: the current state of play. *Health policy*. 1996;37(1):53–72.

8. Propp R, McAdam L, Davis AM, Salbach NM, Weir S, Encisa C, et al. Development and content validation of the muscular dystrophy child health index of life with disabilities questionnaire for children with Duchenne muscular dystrophy. *Dev Med Child Neurol*. 2019;61(1):75–81.
9. Bann CM, Abresch RT, Biesecker B, Conway KC, Heatwole C, Peay H, et al. Measuring quality of life in muscular dystrophy. *Neurology*. 2015; 84(10):1034–42.
10. Wei Y, Speechley K, Campbell C. Health-related quality of life in children with Duchenne muscular dystrophy: a review. *J Neuromuscul Dis*. 2015;2(3): 313–24.
11. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–57.
12. Terwee CB, Prinsen CA, Chiarotto A, Westerman M, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018; 27(5):1159–70.
13. Terwee CB, Prinsen C, Chiarotto A, de Vet H, Bouter LM, Alonso J, et al. COSMIN methodology for assessing the content validity of PROMs—user manual 2018 [Available from: <https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf>]. Accessed 12 July 2018.
14. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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(7):737–45.
15. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *J Clin Epidemiol*. 2018;95:73–93.
16. Mokkink L, Prinsen C, Patrick D, Alonso J, Bouter L, De Vet H, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs)—user manual 2018 [Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf]. Accessed 12 July 2018.
17. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a “Core outcome set”—a practical guideline. *Trials*. 2016;17(1):449.
18. Powell P, Carlton J, Buckley Woods H, Mazzone P. Patient reported outcome measures of quality of life in Duchenne muscular dystrophy (DMD): a systematic review of content and structural validity using COSMIN 2018 [Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018093062]. Accessed 12/03/2019.
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–9.
20. Mackintosh A, Comabella C, Hadi M, Gibbons E, Roberts N, Fitzpatrick R. PROM group construct and instrument type filters 2010 [Available from: <https://www.cosmin.nl/wp-content/uploads/prom-search-filter-oxford-2010.pdf>]. Accessed 11/04/2018.
21. Uttley 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(1):237.
22. Terwee C, Prinsen C, Garotti MR, Suman A, De Vet H, Mokkink L. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res*. 2016;25(4):767–79.
23. COSMIN Group. Search Filters COSMIN 2018 [Available from: <https://www.cosmin.nl/tools/pubmed-search-filters/>]. Accessed 18/09/2018.
24. Mokkink LB, De Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171–9.
25. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651–7.
26. Powell P, Carlton J, Rowen D, Uttley L, Woods HB, Brazier J. Quality of life in Duchenne muscular dystrophy (DMD): initial findings and protocol for a new preference-based measure. *Quality of life research*. 2018;27(Suppl. 1):S170.
27. Powell P, Carlton J. “Social interaction (...) without that then I would probably be miserable”: Understanding quality of life in Duchenne muscular dystrophy. *Quality of life research*. 2019;28(Suppl. 1):S109.
28. Ravens-Sieberer U, Erhart M, Wille N, Wetzel R, Nickel J, Bullinger M. Generic health-related quality-of-life assessment in children and adolescents. *Pharmacoeconomics*. 2006;24(12):1199–220.
29. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj*. 2008;336(7650):924–6.
30. Kohler M, Clarenbach CF, Böni L, Brack T, Russi EW, Bloch KE. Quality of life, physical disability, and respiratory impairment in Duchenne muscular dystrophy. *Am J Respir Crit Care Med*. 2005;172(8):1032–6.
31. Simonds A, Muntoni F, Heather S, Fielding S. Impact of nasal ventilation on survival in hypercapnic Duchenne muscular dystrophy. *Thorax*. 1998;53(11): 949–52.
32. Catalano A, Vita G, Russo M, Vita G, Lasco A, Morabito N, et al. Effects of teriparatide on bone mineral density and quality of life in Duchenne muscular dystrophy related osteoporosis: a case report. *Osteoporos Int*. 2016;27(12):3655–9.
33. Lue Y-J, Chen S-S, Lu Y-M. Quality of life of patients with Duchenne muscular dystrophy: from adolescence to young men. *Disabil Rehabil*. 2017; 39(14):1408–13.
34. Elsenbruch S, Schmid J, Lutz S, Geers B, Schara U. Self-reported quality of life and depressive symptoms in children, adolescents, and adults with Duchenne muscular dystrophy: a cross-sectional survey study. *Neuropediatrics*. 2013;44(05):257–64.
35. McDonald CM, Henricson EK, Abresch RT, Han JJ, Escolar DM, Florence JM, et al. 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.
36. Pangalila RF, Van Den Bos GA, Bartels B, Bergen M, Stam HJ, Roebroek ME. Prevalence of fatigue, pain, and affective disorders in adults with Duchenne muscular dystrophy and their associations with quality of life. *Arch Phys Med Rehabil*. 2015;96(7):1242–7.
37. Pangalila RF, Van Den Bos GA, Bartels B, Bergen MP, Kampelmacher MJ, Stam HJ, et al. Quality of life of adult men with Duchenne muscular dystrophy in the Netherlands: implications for care. *J Rehabil Med*. 2015; 47(2):161–6.
38. Leung D, Herzka D, Thompson R, He B, Bibat G, Tennekoon G, et al. Sildenafil does not improve cardiomyopathy in Duchenne/Becker muscular dystrophy. *Ann Neurol*. 2014;76(4):541–9.
39. Moura MC, Wutzki HC, Voos MC, Resende MBD, Reed UC, Hasue RH. Is functional dependence of Duchenne muscular dystrophy patients determinant of the quality of life and burden of their caregivers? *Arq Neuropsiquiatr*. 2015;73(1):52–7.
40. Zebracki K, Drotar D. Pain and activity limitations in children with Duchenne or Becker muscular dystrophy. *Dev Med Child Neurol*. 2008;50(7):546–52.
41. Bray P, Bundy AC, Ryan MM, North KN, Burns J. Health status of boys with Duchenne muscular dystrophy: a parent’s perspective. *J Paediatr Child Health*. 2011;47(8):557–62.
42. Baiardini I, Minetti C, Bonifacio S, Porcu A, Klersy C, Petralia P, et al. Quality of life in Duchenne muscular dystrophy: the subjective impact on children and parents. *J Child Neuro*. 2011;26(6):707–13.
43. Bendixen RM, Senesac C, Lott DJ, Vandeborne K. Participation and quality of life in children with Duchenne muscular dystrophy using the international classification of functioning, disability, and health. *Health Qual Life Outcomes*. 2012;10(1):43.
44. Bendixen RM, Lott DJ, Senesac C, Mathur S, Vandeborne K. Participation in daily life activities and its relationship to strength and functional measures in boys with Duchenne muscular dystrophy. *Disabil Rehabil*. 2014;36(22):1918–23.
45. Beenakker EA, Fock JM, Van Tol MJ, Maurits NM, Koopman HM, Brouwer OF, et al. Intermittent prednisone therapy in Duchenne muscular dystrophy: a randomized controlled trial. *Arch Neurol*. 2005;62(1):128–32.
46. Cavazza M, Kodra Y, Armeni P, De Santis M, López-Bastida J, Linertová R, et al. Social/economic costs and health-related quality of life in patients with Duchenne muscular dystrophy in Europe. *Eur J Health Econ*. 2016; 17(1):19–29.
47. Pangalila RF, van den Bos GA, Stam HJ, van Exel NJA, Brouwer WB, Roebroek ME. Subjective caregiver burden of parents of adults with Duchenne muscular dystrophy. *Disabil Rehabil*. 2012;34(12):988–96.
48. Graham CD, Rose MR. What explains high life satisfaction in men living with Duchenne muscular dystrophy? A preliminary study to inform psychological intervention. *Muscle Nerve*. 2017;56(1):163–6.

49. Landfeldt E, Lindgren P, Bell CF, Schmitt C, Guglieri M, Straub V, et al. The burden of Duchenne muscular dystrophy: an international, cross-sectional study. *Neurology*. 2014;83(6):529–36.
50. Landfeldt E, Lindgren P, Bell CF, Guglieri M, Straub V, Lochmüller H, 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):508–15.
51. Otto C, Steffensen BF, Højberg A-L, Barkmann C, Rahbek J, Ravens-Sieberer U, et al. Predictors of health-related quality of life in boys with Duchenne muscular dystrophy from six European countries. *J Neurol*. 2017;264(4):709–23.
52. Zamani G, Heidari M, Malamiri RA, Ashrafi MR, Mohammadi M, Badv RS, et al. The quality of life in boys with Duchenne muscular dystrophy. *Neuromuscul Disord*. 2016;26(7):423–7.
53. Opstal SH, Jansen M, van Alfen N, de Groot IJ. Health-related quality of life and its relation to disease severity in boys with Duchenne muscular dystrophy: satisfied boys, worrying parents—a case-control study. *J Child Neurol*. 2014;29(11):1486–95.
54. Reid DT, Renwick RM. Preliminary validation of a new instrument to measure life satisfaction in adolescents with neuromuscular disorders. *Int J Rehabil Res*. 1994;17(2):184.
55. Simon VA, Resende MBD, Simon MA, Zanoteli E, Reed UC. Duchenne muscular dystrophy: quality of life among 95 patients evaluated using the life satisfaction index for adolescents. *Arq Neuropsiquiatr*. 2011;69(1):19–22.
56. Simon VA, Zanoteli E, Simon MAVP, Resende MB, Reed UC. Translation and validation of the life satisfaction index for adolescents scale with neuromuscular disorders: LSI-A Brazil. *Arq Neuropsiquiatr*. 2017;75(8):553–62.
57. Reid D, Renwick R. Relating familial stress to the psychosocial adjustment of adolescents with Duchenne muscular dystrophy. *Int J Rehabil Res*. 2001;24(2):83–93.
58. McDonald CM, McDonald DA, Bagley A, Sienko Thomas S, Buckon CE, Henricson E, et al. Relationship between clinical outcome measures and parent proxy reports of health-related quality of life in ambulatory children with Duchenne muscular dystrophy. *J Child Neurol*. 2010;25(9):1130–44.
59. McDonald CM, Henricson EK, Abresch RT, Duong T, Joyce NC, Hu F, 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):451–61.
60. Ruiz-Cortés X, Ortiz-Corredor F, Mendoza-Pulido C. Reliability of home-based, motor function measure in hereditary neuromuscular diseases. *J Int Med Res*. 2017;45(1):261–71.
61. Henricson E, Abresch R, Han JJ, Nicorici A, Keller EG, de Bie E, et al. The 6-minute walk test and person-reported outcomes in boys with Duchenne muscular dystrophy and typically developing controls: longitudinal comparisons and clinically-meaningful changes over one year. *PLoS Curr*. 2013;5:ecurrents.md.9e17658b007eb79fcd6f723089f79e06.
62. Uzark K, King E, Cripe L, Spicer R, Sage J, Kinnett K, et al. Health-related quality of life in children and adolescents with Duchenne muscular dystrophy. *Pediatrics*. 2012;130(6):e1559–e66.
63. 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(7):879–86.
64. Messina S, Vita GL, Sframeli M, Mondello S, Mazzone E, D'Amico A, et al. Health-related quality of life and functional changes in DMD: a 12-month longitudinal cohort study. *Neuromuscul Disord*. 2016;26(3):189–96.
65. Davis SE, Hynan LS, Limbers CA, Andersen CM, Greene MC, Varni JW, 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(3):97–109.
66. 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):367–73.
67. Hu J, Jiang L, Hong S, Cheng L, Kong M, Ye Y. Reliability and validity of the Chinese version of the pediatric quality of life inventory TM (PedsQL TM) 3.0 neuromuscular module in children with Duchenne muscular dystrophy. *Health Qual Life Outcomes*. 2013;11(1):47.
68. Hu J, Kong M, Ye Y, Hong S, Cheng L, Jiang L. Serum miR-206 and other muscle-specific micro RNA s as non-invasive biomarkers for Duchenne muscular dystrophy. *J Neurochem*. 2014;129(5):877–83.
69. Hu J, Ye Y, Kong M, Hong S, Cheng L, Wang Q, et al. Daily prednisone treatment in Duchenne muscular dystrophy in Southwest China. *Muscle Nerve*. 2015;52(6):1001–7.
70. Bray P, Bundy AC, Ryan MM, North KN, Everett A. Health-related quality of life in boys with Duchenne muscular dystrophy: agreement between parents and their sons. *J Child Neurol*. 2010;25(10):1188–94.
71. Bray P, Bundy AC, Ryan MM, North KN. Can in-the-moment diary methods measure health-related quality of life in Duchenne muscular dystrophy? *Qual Life Res*. 2017;26(5):1145–52.
72. 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.
73. McDonald CM, Henricson EK, Abresch RT, Florence J, Eagle M, Gappmaier E, et al. The 6-minute walk test and other clinical endpoints in Duchenne muscular dystrophy: reliability, concurrent validity, and minimal clinically important differences from a multicenter study. *Muscle Nerve*. 2013;48(3):357–68.
74. Ozyurt G, Bayram E, Karaoglu P, Hiz Kurul S, Yis U. Quality of life and sleep in children diagnosed with Duchenne muscular dystrophy and their mothers' level of anxiety: a case-control study. *Dusunen Adam. J Psychiatry Neurol Sci*. 2015;28(4):362.
75. Schreiber-Katz O, Klug C, Thiele S, Schorling E, Zowe J, Reilich P, et al. Comparative cost of illness analysis and assessment of health care burden of Duchenne and Becker muscular dystrophies in Germany. *Orphanet J Rare Dis*. 2014;9(1):210.
76. Wei Y, Speechley KN, Zou G, Campbell C. The relationship between quality of life and health-related quality of life in young males with Duchenne muscular dystrophy. *Dev Med Child Neurol*. 2017;59(11):1152–7.
77. Teoh LJ, Geelhoed EA, Bayley K, Leonard H, Laing NG. Health care utilization and costs for children and adults with Duchenne muscular dystrophy. *Muscle Nerve*. 2016;53(6):877–84.
78. Sienko S, Buckon C, Fowler E, Bagley A, Staudt L, Sison-Williamson M, et al. Prednisone and deflazacort in Duchenne muscular dystrophy: do they play a different role in child behavior and perceived quality of life? *PLoS Curr*. 2016;8:ecurrents.md.7628d9c014bfa29f821a5cd19723bbaa.
79. Colombo P, Nobile M, Tesei A, Civati F, Gandossini S, Mani E, et al. Assessing mental health in boys with Duchenne muscular dystrophy: emotional, behavioural and neurodevelopmental profile in an Italian clinical sample. *Eur J Paediatr Neurol*. 2017;21(4):639–47.
80. Garralda M, McConachie H, Le Couteur A, Srianjan S, Chakrabarti I, Cirak S, et al. Emotional impact of genetic trials in progressive paediatric disorders: a dose-ranging exon-skipping trial in Duchenne muscular dystrophy. *Child Care Health Dev*. 2013;39(3):449–55.
81. Darke J, Bushby K, Le Couteur A, McConachie H. Survey of behaviour problems in children with neuromuscular diseases. *Eur J Paediatr Neurol*. 2006;10(3):129–34.
82. Orcesi S, Ariaudo G, Mercuri E, Beghi E, Rezzani C, Balottin U. A new self-report quality of life questionnaire for children with neuromuscular disorders: presentation of the instrument, rationale for its development, and some preliminary results. *J Child Neurol*. 2014;29(2):167–81.
83. Grootenhuys MA, De Boone J, Van der Kooij AJ. Living with muscular dystrophy: health related quality of life consequences for children and adults. *Health Qual Life Outcomes*. 2007;5(1):31.
84. Manser RL, Irving L, Byrnes G, Abramson MJ, Stone C, Campbell DA. Screening for lung cancer: a systematic review and meta-analysis of controlled trials. *Thorax*. 2003;58(9):784–9.
85. Mateen F, Oh J, Tergas A, Bhayani N, Kamdar B. Titles versus titles and abstracts for initial screening of articles for systematic reviews. *Clin Epidemiol*. 2013;5:89–95.
86. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961;4(6):561–71.
87. Palermo TM, Witherspoon D, Valenzuela D, Drotar DD. Development and validation of the child activity limitations interview: a measure of pain-related functional impairment in school-age children and adolescents. *Pain*. 2004;109(3):461–70.
88. Petersen C, Schmidt S, Power M, Group ED. Developing and pilot testing a chronic generic quality of life questionnaire for children with chronic conditions. *Qual Life Res*. 2005;7(3):1065–75.
89. Ravens-Sieberer U, Schmidt S, Gosch A, Erhart M, Petersen C, Bullinger M. Measuring subjective health in children and adolescents: results of the European KIDSCREEN/DISABKIDS project. *GMS Psycho Soc Med*. 2007;4:Doc08.

90. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol*. 1989;46(10):1121–3.
91. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166(10):1092–7.
92. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361–70.
93. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. *Pharmacoeconomics*. 1995;7(6):490–502.
94. 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(7):702–22.
95. Vincent K, Carr A, Walburn J, Scott D, Rose M. Construction and validation of a quality of life questionnaire for neuromuscular disease (INQoL). *Neurology*. 2007;68(13):1051–7.
96. Ravens-Sieberer U, Gosch A, Abel T, Auquier P, Bellach B-M, Bruil J, et al. Quality of life in children and adolescents: a European public health perspective. *Sozial-und Präventivmedizin*. 2001;46(5):294–302.
97. Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Duer W, et al. KIDS CREEN-52 quality-of-life measure for children and adolescents. *Expert Rev Pharmacoecon Outcomes Res*. 2005;5(3):353–64.
98. Detmar S, Bruil J, Ravens-Sieberer U, Gosch A, Bisegger C. The use of focus groups in the development of the KIDSCREEN HRQL questionnaire. *Qual Life Res*. 2006;15(8):1345–53.
99. Ravens-Sieberer U, Gosch A, Erhart M, Rueden U, Nickel J, Kurth B-M, et al. The KIDSCREEN questionnaires—quality of life questionnaires for children and adolescents—handbook. Pabst, Wolfgang Science: Lengerich, Germany; 2006.
100. Propp R. Development and psychometric evaluation of the muscular dystrophy child health index of life with disabilities (MDCHILD) questionnaire in children with Duchenne muscular dystrophy; 2017.
101. Varni JW, Burwinkle TM, Katz ER, Meeske K, Dickinson P. The PedsQL™ in pediatric cancer: reliability and validity of the pediatric quality of life inventory™ generic core scales, multidimensional fatigue scale, and cancer module. *Cancer*. 2002;94(7):2090–106.
102. Iannaccone ST, Hynan LS, Morton A, Buchanan R, Limbers CA, Varni JW. The PedsQL™ in pediatric patients with spinal muscular atrophy: feasibility, reliability, and validity of the pediatric quality of life inventory™ generic core scales and neuromuscular module. *Neuromuscul Disord*. 2009;19(12):805–12.
103. Varni JW, Seid M, Rode CA. The PedsQL™: measurement model for the pediatric quality of life inventory. *Med Care*. 1999;37(2):126–39.
104. Spitzer RL, Kroenke K, Williams JB. Patient health questionnaire primary care study group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Jama*. 1999;282(18):1737–44.
105. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–13.
106. Daltroy LH, Liang MH, Fossel AH, Goldberg MJ. The POSNA pediatric musculoskeletal functional health questionnaire: report on reliability, validity, and sensitivity to change. *J Pediatr Orthop*. 1998;18(5):561–71.
107. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res*. 1989;28(2):193–213.
108. Goodman R. The strengths and difficulties questionnaire: a research note. *J Child Psychol Psychiatry*. 1997;38(5):581–6.
109. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. conceptual framework and item selection. *Med Care*. 1992;30(6):473–83.
110. Hays RD, Sherbourne CD, Mazel RM. The rand 36-item health survey 1.0. *Health Econ*. 1993;2(3):217–27.
111. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. *J Epidemiol Community Health*. 1999; 53(1):46–50.
112. Ware JE Jr. SF-36 health survey update. *Spine*. 2000;25(24):3130–9.
113. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. *J Pers Assess*. 1985;49(1):71–5.
114. WHOQOL Group. Development of the WHOQOL: rationale and current status. *Int J Ment Health*. 1994;23(3):24–56.
115. WHOQOL Group. The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. *Soc Sci Med*. 1995;41(10):1403–9.
116. Skevington SM, Arthur PM, Somerset M. Developing items for the WHOQOL: an investigation of contemporary beliefs about quality of life related to health in Britain. *Br J Health Psychol*. 1997;2(1):55–72.
117. WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med*. 1998;28(3):551–8.
118. Chiarotto A, Terwee CB, Kamper SJ, Boers M, Ostelo RW. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain: a systematic review. *J Clin Epidemiol*. 2018;102:23–37.
119. Wiering B, de Boer D, Delnoij D. Patient involvement in the development of patient-reported outcome measures: a scoping review. *Health Expect*. 2017; 20(1):11–23.
120. Carlton J, Peasgood T, Khan S, Barber R, Bostock J, Keetharuth A. An emerging framework for fully incorporating public involvement (PI) into patient-reported outcome measures (PROMs). *J Patient Rep Outcomes*. 2020;4(1):4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

