Measurement properties of multidimensional patient-reported outcome measures in neurodisability: a systematic review of evaluation studies

ASTRID JANSSENS | MORWENNA ROGERS | REBECCA GUMM | CRISPIN JENKINSON | ALAN TENNANT | STUART LOGAN | CHRISTOPHER MORRIS

© 2015 The Authors. Developmental Medicine & Child Neurology published by John Wiley & Sons Ltd on behalf of Mac Keith Press. DOI: 10.1111/dmcn.12982

AIM To identify and appraise the quality of studies that primarily assessed the measurement properties of English language versions of multidimensional patient-reported outcome measures (PROMs) when evaluated with children with neurodisability, and to summarize this evidence.

METHOD MEDLINE, Embase, PsycINFO, CINAHL, AMED, and the National Health Service Economic Evaluation Database were searched. The methodological quality of the papers was assessed using the COSMIN (COnsensus-based Standards for selection of health Measurement Instruments) checklist. Evidence of content validity, construct validity, internal consistency, test–retest reliability, proxy reliability, responsiveness, and precision was extracted and judged against standardized reference criteria.

RESULTS We identified 48 studies of mostly fair to good methodological quality: 37 papers for seven generic PROMs (CHIP, CHQ, CCoL, KIDSSCREEN, PedsQL, SLSS, and YQOL), seven papers for two chronic–generic PROMs (DISABKIDS and Neuro-QOL), and four papers for three preference-based measures (HUI, EQ-5D-Y, and CHSCS-PS).

INTERPRETATION On the basis of this appraisal, the DISABKIDS appears to have more supportive evidence in samples of children with neurodisability. The overall lack of evidence for responsiveness and measurement error is a concern when using these instruments to measure change, or to interpret the findings of studies in which these PROMs have been used to assess change.

Patient-reported outcome measures (PROMs) assess a patient’s health at a single point in time, and are collected through short, self-completed questionnaires. PROMs are advocated for use in clinical trials;1,2 they are also proposed as key performance indicators for evaluating health systems.3 Some PROMs are domain-specific, focusing on a particular aspect of health, such as behaviour; other instruments are multidimensional with sub-scales that assess various aspects of health and wellbeing. PROMs can be condition-specific, designed for use with people with a particular health problem; or they can be generic and therefore appropriate for anyone to report their health; or chronic–generic, designed for people with any long-term health conditions. Preference-based measures incorporate a weighting of scores based on a reference valuation of health states into a single index score; they are used in economic evaluations to assessing cost-effectiveness.2

‘Neurodisability’ is an umbrella term commonly used in the UK for a range of functional problems of neurological origin. Previously, we proposed a definition of neurodisability for children that was supported by many professionals and parents, and indicated a similar grouping of conditions in other countries, albeit with different terminology.4 For some applications in neurodisability, a condition-specific PROM may be preferable if available; for instance, condition-specific measures exist for cerebral palsy (CP)5 and epilepsy.6 However, it is also common for generic PROMs to be utilized in neurodisability, especially for comparison across conditions or with normative samples. Individually, many conditions that result in a neurodisability are rare, but when grouped together they are common. Hence there are situations when it will be expedient for children with neurodisability to be grouped for research, service evaluation, or audits.

When selecting PROMs for a specific purpose, it is necessary to examine both the construct that is being assessed and the measurement properties of candidate instruments.7 Mapping of items in PROMs using the International Classification of Functioning, Disability and Health for Children and Youth is useful to understand the content.
assessed by questionnaires.8,9 Measurement properties should, ideally, have been evaluated in samples representative of the intended population to determine whether the instrument is applicable for that population.1 Language and cultural issues also affect how people interpret and respond to questions; hence one cannot simply assume that PROMs perform consistently across languages and cultures.10,11 Therefore, the US Food and Drug Administration recommends that evidence be provided of the process used to test measurement properties in the language where assessments will be made.1

Scale development methodology has evolved in recent years and approaches using item response theory are more commonly utilized; such approaches use mathematical models to examine responses to individual items in questionnaires and offer better scale precision.12 Methods for appraising the evidence of psychometric performance on measures have also become more standardized.13 The COnsensus-based Standards for selection of health Measurement INstruments (COSMIN) was developed to enable a standardized assessment of the methodological quality of research studies evaluating measurement properties of tools.13,14

In previous papers we documented a systematic review of generic multidimensional PROMs for children and young people, in which we mapped the content to the International Classification of Functioning, Disability and Health for Children and Youth insofar as it was possible,9 and appraised studies evaluating measurement properties in general population samples.15 In this study we build upon that foundation by focusing on evaluations of the PROMs identified in the previous systematic review where they have been tested in samples of populations with neurodisabilities. In this instance, as the aim was to examine which instruments could be considered robust for application across children with neurodisability, we also included chronic–generic tools. We use the word ‘PROM’ to refer to the group of questionnaires (different versions according to age group, length, or responder) of a certain instrument; we use the word ‘questionnaire’ to refer to a specific version of an instrument. A list of the PROMs, the different types of questionnaires, and full names is presented in Table I.

### METHOD

#### Search strategy

Candidate generic PROMs were identified and catalogued as part of a previous systematic review.9 In addition, chronic–generic PROMs were included as they could be used across neurodisability conditions; three eligible chronic–generic tools were identified in our broader research programme (Disabkids, Functional Disability Index, Neuro-QoL). For this review, three groups of search terms were combined: the names of the PROMs and their synonyms; terms for children; and terms for neurodisability, for which both free text and medical subject headings were used.

Searches were conducted on MEDLINE (including in-process and other non-indexed citations), Embase, PsycINFO, and AMED (via OvidSP), CINAHL (via EBSCO-host), and the National Health Service Economic

---

### Table I: PROMs (group of questionnaires), the different versions (according to age group, length, or responder), and acronyms

| Overall PROM name | Acronym questionnaire | Full name questionnaire |
|-------------------|-----------------------|-------------------------|
| CHIP              | CHIP-CE CRF           | Child Health And Illness Profile – Child Edition Child Report Form |
|                   | CHIP-CE PRF           | Child Health And Illness Profile – Child Edition Parent Report Form (45-item) |
|                   | CHIP-CE PRF           | Child Health And Illness Profile – Child Edition Parent Report Form (76-item) |
| CHQ               | CHQ-PF28              | Child Health Questionnaire Parent Short Form |
|                   | CHQ-PF50              | Child Health Questionnaire Parent Long Form |
|                   | CHQ-CF87              | Child Health Questionnaire Child Form (87-item) |
| CHSCS-PS          | CHSCS-PS              | Comprehensive Health Status Classification System – Preschool |
| CQol              | CQol                  | Child Quality of Life Questionnaire |
| DISABKIDS         | DISABKIDS DGM-37      | DISABKIDS Chronic Generic Measure – long form |
|                   | DISABKIDS Smileys-6   | DISABKIDS Smile Measure |
| EQ-5D-Y           | EQ-5D-Y               | EuroQol 5D Youth |
| HUI                | HUI                   | Health Utilities Index 2 |
|                   | HUI2                  | Health Utilities Index 3 |
| KIDSCREEN         | KIDSCREEN-52          | KIDSCREEN-52 |
|                   | KIDSCREEN-10          | KIDSCREEN-10 |
| Neuro-QOL         | Neuro-QOL             | Neurology Quality of Life Measurement System |
| PedoQL            | PedoQL Infant Scales  | Pediatric Quality Of Life Inventory Trade Mark 4.0 – Infant Scales |
|                   | PedoQL                | Pediatric Quality Of Life Inventory Trade Mark 4.0 – Generic Core Scales |
|                   | PedoQL SF15 GC        | Pediatric Quality Of Life Inventory Trade Mark 4.0 – Short Form 15 |
| SLSS              | SLSS                  | Student Life Satisfaction Scale |
|                   | BMSLSS                | Brief Multi-dimensional Student Life Satisfaction Scale |
| YQoL              | YQoL-S                | Youth Quality Of Life instrument – Surveillance version |
|                   | YQoL-R                | Youth Quality Of Life instrument – Research version |
Evaluation Database (NHS EED; via the Cochrane Library). No date restriction was applied. The searches were run between 12 and 25 September 2012 and updated on 30 July 2014. Forward (checking if key papers had been cited) and backward (checking reference lists) citation chasing was performed for key references to ensure that all relevant literature was retrieved. The electronic search strategy designed for MEDLINE and translated for the other databases is presented in Appendix S1 (online supporting information).

Inclusion and exclusion criteria
Articles were selected when written in English and reporting on a study that was (1) specifically designed to evaluate the psychometric properties of candidate PROMs using an English language version of the questionnaire, (2) conducted in a population including at least 10% children up to 18 years old with neurodisability, or mixed chronic conditions including neurodisability, and (3) published in a peer-reviewed journal. Articles were excluded if (1) the PROM was used as a criterion standard to test another instrument, (2) less than 10% of the study population was younger than 18 years, (3) less than 10% of the study sample was diagnosed with neurodisability.

Study selection
Titles and abstracts of all unique citations were screened against the eligibility criteria by two reviewers (AJ and RG/CM); any disagreements were resolved by discussion. The full text of any potentially relevant article was retrieved and screened using the same procedure. A flowchart describing the process of study selection can be found in Figure 1.

Assessment of methodological quality of included articles
For each included paper, the COSMIN checklist was used to appraise the methodological quality of the study and the completeness of the report. We assessed the methods and reporting of how the following properties had been tested: internal consistency, reliability, measurement error, content validity, structural validity, hypothesis testing, criterion validity, and responsiveness. Cross-cultural validity was not examined as the purpose of the work was to inform UK health services policy where currently only English language versions are administered. The COSMIN checklist uses a ‘worst score counts’ rating of methodological quality as excellent, good, fair, or poor based on factors such as adequate sample size and appropriate statistical methods used. The checklist was administered by two reviewers (AJ and CM); discrepancies were resolved by discussion.

Data extraction
For each included paper the following descriptive data were extracted using a standardized, piloted data extraction form: first author name and year, name and version of the instrument (including child or parent version), study aim, study population (participants’ characteristics including type of neurodisability and diagnosis), number of participants, age range, mean age (and standard deviation), and setting or country where the study was conducted. Data were extracted by one reviewer (RG/AJ) and a 10% sample was checked by a second (AJ/CM).

Then, any data on evidence of the measurement properties of instruments were extracted including content validity (theoretical framework and/or qualitative research), construct validity (structural validity concerning how domain sub-scales were determined for instance using factor analysis, and hypothesis testing to verify sub-scales measure the intended construct), internal consistency (including domain sub-scales where appropriate), test–retest reliability, proxy reliability (between child and parent), precision, and responsiveness (whether the increases/decreases in scores can be considered robust and exceed measurement error). Data were extracted by one reviewer (RG/AJ) and checked by a second (AJ/CM); disagreements were resolved by discussion.

Appraisal of measurement properties and summary of evidence
Evidence of measurement properties was judged using standardized reference criteria and thresholds (Table II). These data were summarized in a single rating for each measurement property following methods commonly used for the presentation of such findings. To summarize available evidence, we took into account the following elements (Table III): (1) data extracted from included studies, with reference to standard criteria (Table II); (2) the methodological quality of studies (COSMIN) and number of studies; and (3) the thoroughness of testing, giving further weight to any studies that appeared not to have been conducted by the original developers. Two reviewers (AJ and CM) made the judgement through discussion based on available evidence.

RESULTS
We found 48 papers that report evaluations of measurement properties of 12 PROMs (see Table IV): 37 papers for seven generic PROMs (CHIP, CHQ, CQoL, KIDSCREEN, PedsQL, SLSS and YQOL), seven papers for two chronic–generic PROMs (DISABKIDS and Neuro QOL), and four papers for three preference-based measures (HUI, EQ-5D-Y and CHSCS-PS) (Table IV). Twenty papers described evaluations of the PedsQL4.0 in various neurodisability samples, and 11 papers pertained to various versions of the CHQ. The most common conditions in samples were CP, epilepsy, attention-deficit–hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and traumatic brain injury. The evaluations spanned children in a variety of age groups from 2 to 18 years old. The evaluations were performed in Canada, USA, Europe, and Australia.

The methodological quality of the included studies was variable (Table V). Internal consistency, test–retest reliability, and construct validity (hypothesis testing) have been more frequently studied in neurodisability samples; several studies have examined structural validity; very few studies...
have evaluated responsiveness and measurement error. A summary appraisal of the evidence for measurement properties of each PROM is given in Table VI.

The PedsQL has been evaluated with children with a wide range of neurodisability including CP, ADHD, ASD, acquired brain injury, neuromuscular, and neuro-oncology conditions. Although there is supportive evidence for the structural validity and test–retest reliability of the PedsQL, there is conflicting evidence for the internal consistency of the subscales, particularly the school functioning domain.
Evidence not in favor

Reproducibility Test
- retest reliability: intraclass correlation coefficient > 0.7 adequate, > 0.9 excellent.
- Proxy-reliability: child- and parent-reported reliability intraclass correlation coefficient > 0.7

Internal consistency
- Cronbach's alpha coefficient > 0.7 and < 0.9

Precision
- Assessment of measurement error; floor or ceiling effects < 15%; evidence provided by Rasch analysis and/or interval level scaling

Responsiveness
- Longitudinal data about change in scores with reference to hypotheses, measurement error, and minimal important difference

| Table II: Appraisal of psychometric properties and indicative criteria |
|---------------------------------------------|
| Psychometric property | Indicative criteria |
|---------------------------------------------|
| **Content validity** | Clear conceptual framework consistent with stated purpose of measurement. Qualitative research with potential respondents |
| **Construct validity** | Structural validity from factor analysis. Post-hoc tests of unidimensionality by Rasch analysis. Hypothesis testing, with a priori hypotheses about direction and magnitude of expected effect sizes. Tests for differential item and scale functioning between sex, age groups, and different diagnoses |
| **Reproducibility** | Test-retest reliability: intraclass correlation coefficient > 0.7 adequate, > 0.9 excellent. Proxy-reliability: child- and parent-reported reliability intraclass correlation coefficient > 0.7 |
| **Internal consistency** | Cronbach’s alpha coefficient > 0.7 and < 0.9 |
| **Precision** | Assessment of measurement error; floor or ceiling effects < 15%; evidence provided by Rasch analysis and/or interval level scaling |
| **Responsiveness** | Longitudinal data about change in scores with reference to hypotheses, measurement error, and minimal important difference |

| Table III: Indices for summarizing appraising psychometric properties of patient-reported outcome measures |
|---------------------------------------------|
| Rating | Definition |
|---------------------------------------------|
| 0 | Not reported |
| 7 | Not clearly determined |
| - | Evidence not in favour |
| +/- | Conflicting evidence |
| + | Some evidence in favour |
| ++ | Some good evidence in favour |
| +++ | Good evidence in favour |

which scored consistently low (0.45–0.65).21–24 Other papers reported values of Cronbach’s alpha below 0.7 for emotional functioning,25 social functioning,25,26 and physical functioning.22 We found conflicting evidence for precision; overall floor and ceiling effects were less than 15% for most scales, except social functioning (up to 36%).21,27 The responsiveness of the PedsQL was assessed in one poor-quality study, thus a rating was not determined.28 Several studies reported child-proxy reliability, all reporting low to moderate agreement (intraclass correlation coefficient 0.10–0.75, with most between 0.20 and 0.60).

Versions of the CHQ have been evaluated with children with CP, ADHD, acquired brain injury, and epilepsy. There is supportive evidence for the structural validity of child and parent report versions of the CHQ and internal consistency of the child report version. Both parent versions show poor results for ceiling and floor effects on several scales; ceiling effects are found for most of the individual scales, with scores up to 86% for role and social functioning.29 One study reports low values of Cronbach’s alpha for the domains family cohesion, bodily pain role/social functioning, and role/social limitations of the 28-item version;35 three studies report conflicting findings for the 50-item version, with one paper reporting supportive evidence for all domains30 and two studies reporting low alpha scores for general health perceptions and family impact (emotional and time).29,31 We also found conflicting evidence for the CHQ-PF50 for construct validity and responsiveness.

The DISABKIDS was developed for and with children who have chronic health conditions including CP and epilepsy. Supportive evidence from methodologically robust studies exists for content validity, construct validity, and internal consistency of the 37-item version, and there is favourable evidence for structural validity, test–retest reliability, and precision. Evidence did not support child-proxy reliability. For the 6-item version for younger children, evidence supports the content validity, structural validity, test–retest reliability, but is conflicting for internal consistency, with values of Cronbach’s alpha dropping just below 0.70 for the child version.32

Kidscreen-52 has been evaluated in one study using Rasch analysis with data from children with CP in countries across Europe; the findings for the English language version were reported separately.33 Evidence supports structural validity, construct validity, and precision of Kidscreen-52. Supporting evidence was found for internal consistency and test–retest reliability for the 10-item version in one study of poorer quality, including children with CP.34

One methodologically robust study evaluated the SLSS and BMSLSS with adolescents with conditions including CP, acquired brain injury, and ASD, providing evidence for construct validity, structural validity, internal consistency, and test–retest reliability.34

Two papers evaluating the CHIP-CE parent report version with children with ADHD support structural validity, construct validity, internal consistency and precision.35,36

Each of the four preference-based measures has been evaluated in one study. Evidence from hypothesis testing supports the construct validity of the CHSCS-PS and HUI3, but was inconsistent for the EQ-5D-Y. The child-proxy reliability of the HUI2 was not rated as the study was of poor quality.
| Instrument version | Reference | Aim/purpose | Study population | n   | Age range, y | Mean age, y (SD) | Setting, country |
|--------------------|-----------|-------------|------------------|-----|-------------|------------------|-----------------|
| CHIP-CE PR         | Riley35   | To test reliability and validity of the CHIP-CE with children with ADHD | Children with ADHD in a clinical trial | 1476 | 6-18 | Not stated | Outpatient clinics, Europe |
| CHIP-CE PR         | Schacht36 | To test reliability and validity of the CHIP-CE with children with ADHD | Children with ADHD in five clinical trials | 794 | 6-15 | 9.7 (2.30) | Outpatient clinics, Europe and Canada |
| CHO-CF87 PR        | Landgraf59| To test reliability and validity of CHQ-CF87 with ADHD | General population, subgroup of children with ADHD, and children with end stage renal disease | 354 (total) SRFNDP | 9-16 | 11.8 (1.9) | Postal survey, USA |
| CHQ-PF28 PR        | Pencharz20 | Evaluate and compare the psychometric properties of the CHQ-PF-28 in a paediatric clinical sample | CYP with musculoskeletal disorders, including children with CP and DMD | 166 | 5-16 | 11.0 (2.9) | Hospital and paediatric rehabilitation centre, Canada |
| CHQ-PF50 PR        | Vitale61  | Evaluate and compare the psychometric properties in a paediatric orthopaedic sample | Children with a range of musculoskeletal problems, including CP | 242 | 5-18 | 12 | Physician’s office, USA |
| CHQ-PF50 PR        | Wake30    | To test reliability and validity of the CHQ-PF50 with CP | Children with CP | 80 | 5-18 | 11.25 (3.5) | Outpatient clinics, Australia |
| CHQ-PF50 PR        | Rentz29   | To test reliability and validity of the CHQ-PF50 with children with ADHD | Children with ADHD in a clinical trial | 921 | 6-18 | 11 | Outpatient clinics, USA |
| CHQ-PF28 PR        | Vitale62  | To determine the efficacy and sensitivity of the CHQ in children with CP | Children with CP | 180 | 5-18 | 10.7 | Completed before treatment for CP at one hospital, USA Inpatient clinic, Canada |
| CHQ-PF50 PR        | Thomas-Stonell63 | To test responsiveness of the CHQ-PF50 with TBI | Paediatric patients with TBI (1) Children with chronic conditions, including epilepsy (2) General paediatric population | 33 | 4-18 | 12.5 (4.5) | (1) Outpatient clinics, USA (2) Comparison group from sleep study, USA |
| CHQ-PF50 PR        | Drotar64  | To test factor structure of the CHQ-PF-50 in a sample of children and adolescents with chronic conditions and physically healthy children seen in a paediatric setting | (1) Children with chronic conditions, including epilepsy (2) General paediatric population | 661 (1) 329 (total) Epilepsy: 25 (2) 332 | 5-18 | (1) 12.3 (3.5) (2) 11.4 (3.5) | (1) Outpatient clinics, USA (2) Comparison group from sleep study, USA |
| CHQ-PF50 PR        | McCullough31 | To test reliability and validity of CHQ with children with CP | Children with CP | 818 | 8-12 | Not stated | Home visits, Europe |
| CHQ-PF28 PR        | Davis25   | To compare reliability and validity of the CHQ-PF28 and Kidscreen-10 | Children with CP | 204 (PR) 54 (SR) | 4-12 | 8.25 (2.51) | Outpatient clinics, Australia |
| CHO-PF50 PR        | Ferro65   | To investigate higher-order factor structure of the CHQ-PF50 in children with new onset epilepsy | Children newly diagnosed with epilepsy | 374 | 4-12 | 7.4 (2.3) | Paediatric neurologists’ patients, Canada |
| CHSCS-PS PR        | Saigaf66  | To develop a multi-dimensional health status classification system for preschool children | (1a) VLBW children and (1b) general population sample (2) VLBW children (3) Children with CP | (1a) 101 (1b) 50 (2) 150 (3) 222 SRFNDP | 1-6 | (1a) 3.05 (0.09) (1b) 3.04 (0.08) (2) 3.88 (0.62) (3) 3.79 (1.01) | Outpatient clinics, Canada and Australia |
| Reference | Aim/purpose | Study population |
|-----------|-------------|------------------|
| Graham | To develop a QoL measure for 9- to 15-y-old children, and test it in a healthy and three clinical samples | (1) Children with chronic physical disorders, including neurological disorders; children with mental retardation; children with psychiatric disorders (2) General paediatric population |
| Petersen | To develop and test a chronic–generic HRQoL measure CYP with different chronic health conditions | 360 CP: 21 Epilepsy: 37 |
| Schmidt | To test cross-cultural validity of the DISABKIDS in children with different chronic conditions, including CP and epilepsy | 122 CP: 27 Epilepsy: 45 |
| Simeoni | To shorten and test the shortened version of the DISABKIDS in children with chronic diseases | 122 CP: 27 Epilepsy: 45 |
| Chaplin | To test the reliability and validity of the DISABKIDS Smileys in children with a chronic disease | 435 CP: 56 Epilepsy: 40 |
| Matza | To test EQ-5D with children with ADHD, correlations with CHQ-PF50 and CHIP-CE | 126 (total) 83 (UK) |
| Glaser | To assess interrater reliability of the HUI | 30 6–16 |
| Tilford | To evaluate the construct validity of the HUI-3 with children with ASD | 150 4–17 8.6 (3.3) |
| Erhart | To test reliability and validity of KIDSCREEN-52 in children with CP | Various clinics, USA, and six other European countries, UK |
| Perez | To identify content area for a health-related quality of life instrument in neurology | Two focus groups (no numbers stated) 14–20 15.83 (2.23) |
| Cella | To develop and calibrate the Neuro QOL scales | 59 Epilepsy: 50; MD: 9 |
| Lai | To test reliability and dimensionality of the instrument using computerized adaptive testing | 117 Epilepsy: 111 MD: 60 10–21 14.5 (2.8) |

**Table IV:** Continued
Table IV: Continued

| Instrument version | Reference | Aim/purpose | Study population | Setting, country |
|--------------------|-----------|-------------|------------------|-----------------|
| PedQL 4.0 SR and PR | Eiser74 | (1) To test inter-rater reliability of the PedQL in children and adolescents with a CNS tumour and (2) to examine intra-rater reliability of the PedQL in children with a CNS tumour | (1) CYP who has survived a CNS tumour (2) CYP with a history of a CNS tumour | Outpatient clinics, USA |
| PedQL 4.0 SR and PR | McCarthy75 | (1) To test reliability and validity of the PedQL in children and adolescents with TBI | Children and adolescents with TBI | Telephone interviews, USA |
| PedQL 4.0 SR and PR | Varni22 | To test reliability and validity of the PedQL in children with ADHD | Children with ADHD | Outpatient clinics, USA |
| PedQL 4.0 SR and PR | Varni76 | (1) To test reliability and validity of the PedQL in children with CP | Children with CP | Outpatient clinics, USA |
| PedQL 4.0 SR and PR | Varni27 | (1) To test reliability and validity of the PedQL parent-proxy report in children with chronic health conditions, including ADHD and CP | Children with chronic health conditions, including ADHD and CP | Outpatient clinics, USA |
| PedQL 4.0 SR and PR | Majnemer80 | To examine the internal consistency of the self-reported PedQL and the Pediatric Quality of Life Inventory (PedsQL) in children and adolescents with complex physical health conditions, including CP and SMA | Children with complex physical health conditions, including CP and SMA | Outpatient clinics, USA |
| PedQL 4.0 SR and PR | Limbers81 | To examine the feasibility, reliability, and validity of the PedQL in children with Asperger syndrome | Children with Asperger syndrome | Outpatient clinics, USA |
| PedQL 4.0 SR and PR | Davis23 | To examine the feasibility, reliability, and validity of the PedQL in children with DMD | Children with DMD | Outpatient clinics, USA |
| Instrument version<sup>a</sup> | Reference | Aim/purpose | Study population | n  | Age range, y | Mean age, y (SD) | Setting, country |
|-----------------------------|-----------|-------------|------------------|----|--------------|------------------|-----------------|
| PedsQL 4.0 SR and PR | Dunaway<sup>82</sup> | To test reliability of telephone administration | Children with SMA | 20 | 2–18 | 8.4 SD not stated | Outpatient clinics, USA |
| PedsQL 4.0 SR | Limbers<sup>24</sup> | To test reliability and validity of the PedsQL with children with ADHD | Children with ADHD | 183 | 5–18 | 11.08 (3.7) | Outpatient clinics, USA |
| PedsQL 4.0 SR | Shipman<sup>83</sup> | To test reliability and validity of the PedsQL with children with ASD | Children with ASD | 39 | 12–18 | 14.8 SD not stated | Outpatient clinics, USA |
| PedsQL 4.0 SR and PR | Green<sup>84</sup> | To investigate parent-adolescent agreement in long-term QOL outcomes | Adolescents who sustained TBI between birth and 5 y old | 16 | 15–18 | 16.5 (1) | Phone interview, recruited at neurosurgical ward of a children’s hospital, Australia |
| PedsQL 4.0 SR and PR | Shekdrick<sup>95</sup> | To compare adolescent self-reports with parent reports regarding the QOL of adolescents with ASD | Adolescents diagnosed with ASD at a developmental-behavioural clinic | 39 | 12–18 | 14.8 | Recruit at, and procedures completed at clinic, USA |
| PedsQL 4.0 SR and PR | Taverno<sup>86</sup> | To evaluate the content validity of the PedsQL for use with children with ASD | Children diagnosed with ASD | 10 (P) 4 (YP) | 9–12 | Not stated | Recruit via the Database of Children Living with ASDs, UK |
| SLSS and BMSLSS SR and PR | McDougall<sup>84</sup> | To assess to psychometric properties of the BMSLSS and SLSS in youth with chronic conditions | Adolescents with chronic conditions (including CP, acquired brain injury, and ASD) | 439 (Total) CP: 150 (35%) Acquired brain injury: 59 (14%) ASD: 35 (7%) | 11–17 | Not stated | In a treatment office or adolescent’s home, Canada |
| YQoL-R SR | Patrick<sup>87</sup> | To develop a quality of life measure for adolescents | Adolescents including samples of general population, ADHD, and mobility disability | 12–18 | Not stated | | Outpatient clinics, USA |

<sup>a</sup>Definitions of the instruments are presented in Tables I and SI (online supporting information). PR, parent report; ADHD, attention-deficit-hyperactivity disorder; SR, self-report; SRFNDP, separate results reported for reported population with neurodisability; CP, cerebral palsy; DMD, Duchenne muscular dystrophy; CYP, Children and young people; SRFESP, Separate results reported for English speaking population; MD, muscular dystrophy; TBI, traumatic brain injury; VLBW, very low birthweight; QoL, quality of life; HRQoL, health-related quality of life; SMA, spinal muscular atrophy; CNS, central nervous system; ASD, autism spectrum disorder; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.88
Table V: Methodological quality of studies evaluating measurement properties of candidate patient-reported outcome measures in a population with neurodisability

| Instrument version | Author                  | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypothesis testing | Criterion validity | Responsiveness |
|--------------------|-------------------------|----------------------|-------------|-------------------|-----------------|---------------------|--------------------|-------------------|-----------------|
| CHIP-CE            | Riley                   | Good                 | Fair        | Good              |                 |                     | Good               | Good              |                 |
| CHIP-CE            | Schacht                 | Fair                 | Fair        | Good              |                 |                     | Good               | Fair              |                 |
| CHQ-CF87           | Landgraf                | Good                 | Fair        | Good              |                 |                     | Good               | Good              |                 |
| CHQ-PF28           | Pencharz                | Good                 | Fair        | Good              |                 |                     | Good               | Good              |                 |
| CHQ-PF50           | Vitale                  | Good                 | Fair        | Good              |                 |                     | Good               | Good              |                 |
| CHQ-PF50           | Wake                    | Good                 | Fair        | Good              |                 |                     |                    | Good              |                 |
| CHQ-PF50           | Vitale                  | Good                 | Fair        | Good              |                 |                     | Good               | Good              |                 |
| CHQ-PF50           | Rentz                   | Good                 | Fair        | Good              |                 |                     | Good               | Good              |                 |
| CHQ-PF50           | Drotar                  | Good                 | Poor        | Good              |                 |                     | Good               | Good              | Fair            |
| CHQ-PF50           | Thomas-Stonell          | Good                 | Poor        | Good              |                 |                     | Good               | Good              |                 |
| CHQ-PF50           | McCullough              | Excellent            | Poor        | Good              |                 |                     | Good               | Good              |                 |
| CHQ-PF28           | Davis                   | Fair                 | Good        | Fair              |                 |                     | Good               | Good              |                 |
| KIDSCREEN-10       | Ferro                   | Excellent            | Poor        | Poor              |                 |                     | Good               | Good              |                 |
| CHQ-PF50           | Saiga                   | Excellent            | Poor        | Poor              |                 |                     | Good               | Good              |                 |
| CQoL               | Graham                  | Fair                 | Poor        | Fair              |                 |                     | Good               | Good              |                 |
| DCGM-37            | Petersen                | Good                 | Fair        | Good              |                 |                     | Good               | Good              |                 |
| DCGM-37            | Schmidt                 | Good                 | Fair        | Good              |                 |                     | Good               | Good              | Good            |
| DCGM-37            | Simeoni                 | Good                 | Fair        | Good              |                 |                     | Good               | Good              | Good            |
| DISABKIDS Smiley-6 | Chaplin                 | Poor                 | Poor        | Good              |                 |                     | Fair               | Fair              | Fair            |
| EQ-5D-Y            | Matza                   | Fair                 | Poor        | Good              |                 |                     | Fair               | Fair              | Fair            |
| HUI2               | Glaser                  | Poor                 | Fair        | Poor              |                 |                     | Fair               | Fair              |                 |
| HUI3               | Tilford                 | Good                 | Poor        | Poor              |                 |                     | Good               | Good              |                 |
| KIDSCREEN-52       | Erhart                  | Poor                 | Good        | Good              |                 |                     | Good               | Good              |                 |
| Neuro QOL          | Perez                   | Fair                 | Poor        | Good              |                 |                     | Poor               | Good              |                 |
| Neuro QOL          | Cella                   | Poor                 | Good        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Eiser                   | Poor                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | McCarthy                | Good                 | Fair        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Varni                   | Fair                 | Fair        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Varni                   | Fair                 | Good        | Fair              |                 |                     | Fair               | Fair              |                 |
| PedsQL 4.0         | Varni                   | Fair                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Varni                   | Fair                 | Fair        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Palmer                  | Poor                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Majnemer                | Fair                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Oeffinger               | Poor                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Varni                   | Poor                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Young                   | Poor                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Limbers                 | Fair                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Iannaccone              | Fair                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Davis                   | Poor                 | Fair        | Poor              |                 |                     | Fair               | Fair              |                 |
| PedsQL 4.0         | Dunaway                 | Poor                 | Fair        | Poor              |                 |                     | Fair               | Fair              |                 |
| PedsQL 4.0         | Limbers                 | Poor                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Shipman                 | Poor                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Green                   | Poor                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Sheldrick               | Poor                 | Poor        | Poor              |                 |                     | Poor               | Good              |                 |
| PedsQL 4.0         | Tavernor                | Fair                 | Poor        | Poor              |                 |                     | Fair               | Good              |                 |
| SLSS and BMSLSS     | McDougal                | Excellent            | Fair        | Excellent         |                 |                     | Excellent          | Fair              | Fair            |
| YQLQ               | Patrick                 | Fair                 | Poor        | Fair              |                 |                     | Poor               | Good              |                 |

*aDefinitions of the instruments are presented in Tables I and SI (online supporting information).*
The CQoL was developed with children with intellectual disability, chronic physical disorders, and psychiatric disorders, and parents. The study reporting the development and preliminary testing provides supportive evidence for content validity; internal validity and test–retest reliability could not be determined as these elements of the study were of poor quality. The YQOL was developed for and with children with disabilities. However, the study reporting on the content validity of the instrument does not state which conditions were included. A companion paper reports supportive evidence on structural and construct validity, and internal consistency in a study of moderate quality, including children with ADHD or mobility disability.

We identified three papers reporting on the development and initial testing of the Neuro-QOL; two papers report the same data. Epilepsy and muscular dystrophy were selected as conditions for test development of the paediatric Neuro-QOL item pool. The content validity, reported in three papers, was rated as good. Domains were identified through a literature review, expert interviews, parent and carer focus groups, and keyword search. Cognitive interviews were conducted with children aged 10 to 18 years to ensure appropriate understanding and literacy levels. Other measurement properties were not rated owing to the poor quality of the studies.

**DISCUSSION**

This review identified 12 multidimensional PROMs, with 18 versions of questionnaires, that have been evaluated with children with various neurodisability conditions, including CP, ADHD, ASD, epilepsy, acquired brain injury, neuromuscular and neuro-oncology conditions. The PedsQL and CHQ have been evaluated more than other instruments, though some of the evidence undermines confidence in their ability to produce robust measurement. On the basis of this appraisal, the DISABKIDS appears to have more evidence to support its measurement properties in samples of children with neurodisability. None of the PROMs has been evaluated comprehensively across all relevant measurement properties, with responsiveness and measurement error being the least studied.

The paucity of evidence available for the properties of responsiveness and measurement error should be a concern for anyone wishing to use the instruments to measure change, or for those seeking to interpret the findings of studies in which these PROMs have been used to assess change. This gap needs to be evaluated in paediatric populations with neurodisability to inform decisions about what constitutes meaningful change scores. Changes in scores may be statistically significant, especially in large samples, but may not be clinically important. Indices such as the minimal clinically important difference is the mean change in score reported by the respondents who indicate that they had noticed some small change. The minimal clinically important difference has been evaluated for the PedsQL in a sample of children with diabetes; nevertheless one cannot necessarily assume this difference will be the same for children with neurodisability conditions. Other ways to address the lack of evidence for responsiveness include the minimum detectable change, which is an indication of the amount of change required to have confidence that any observed change is beyond measurement error; a common standard is to use a 90% confidence level. The effect size is calculated by dividing the amount of change by the standard deviation of the baseline score. Revicki et al. suggest calculating different indices of minimal change, and for these to triangulate towards a range of values, in which confidence increases with replication.

There appears a dearth of evaluations of the measurement properties of preference-based measures in children with neurodisability, adding to the lack of evidence for these instruments in general populations. Internal consistency may conflict with the underlying theory of health

---

**Table VI: Summary appraisal of measurement properties in a population with neurodisability**

| Instrument version* | Content validity | Structural validity | Construct validity | Internal consistency | Test-retest reliability | Proxy reliability | Precision | Responsiveness |
|---------------------|------------------|---------------------|--------------------|---------------------|------------------------|------------------|----------|---------------|
| BMSLSS              | 0                | ++                  | +                  | +                   | 0                      | 0                | 0        | 0             |
| CHIP CE             | 0                | ++                  | +                  | 0                   | 0                      | 0                | 0        | 0             |
| CHQ-CF87            | 0                | +                   | +                  | 0                   | 0                      | 0                | 0        | 0             |
| CHQ-PF28            | 0                | +                   | 0                  | –                   | 0                      | 0                | 0        | 0             |
| CHQ-PF50            | 0                | +                   | +/–                | +/–                 | 0                      | 0                | –        | +/-           |
| CHQ-52              | 0                | 0                   | +                  | 0                   | 0                      | 0                | 0        | 0             |
| CQoL                | ++               | 0                   | +                  | 0                   | 0                      | 0                | 0        | 0             |
| DCMD-37             | ++               | ++                  | +/–                | +/–                 | 0                      | 0                | 0        | 0             |
| DISABKIDS Smiley-6  | ++               | ++                  | 0                  | +/–                 | 0                      | 0                | 0        | 0             |
| EQ-SD-Y             | 0                | 0                   | +/-                | 0                   | 0                      | 0                | 0        | 0             |
| HU12                | 0                | 0                   | 0                  | 0                   | 0                      | 0                | 0        | 0             |
| HUI3                | 0                | 0                   | +                  | 0                   | 0                      | 0                | 0        | 0             |
| KIDSCREEN-52       | 0                | +                   | ++                 | 0                   | 0                      | 0                | 0        | 0             |
| KIDSCREEN-10       | 0                | 0                   | 0                  | +                   | 0                      | 0                | 0        | 0             |
| Neuro-QOL           | + ?              | +                   | +/–                | +/–                 | 0                      | 0                | 0        | 0             |
| PedsQL 4.0         | 0                | +                   | +/–                | +/–                 | 0                      | 0                | 0        | 0             |
| SLSS                | 0                | +                   | +                  | +                   | 0                      | 0                | 0        | 0             |
| YQoL                | 0                | +                   | +                  | +                   | 0                      | 0                | 0        | 0             |

*Definitions of the instruments are presented in Tables I and SI (online supporting information).
economic instruments, but the properties of face, content and construct validity, and test–retest reliability remain requisite. Lack of evidence for these measurement properties undermines confidence in health economic evaluations based on preference-based measures. We did not examine the methods used to derive the scaling of the preference-based measures; the methods for creating the preference weighting were assumed to produce interval-level measurement. As the purpose of preference-based measures is to quantify the value or strength of preference for health change, the means for assuming and eliciting preference values should be critically assessed.

The information from this review makes it difficult to recommend a multidimensional PROM for use in paediatric neurodisability based on measurement properties established in relevant conditions. Our review of evaluations of generic multidimensional PROMs in general population samples identified 41 potentially eligible PROMs, and identified 126 papers that reported evidence of the measurement properties of 25 PROMs using English-language versions in general population samples. Although robust evidence was lacking for one or more properties for all PROMs, there was evidence to support more measurement properties for the CHIP, Healthy Pathways, KIDSCREEN, and MSLSS. The CHU-9D was the preference-based measure with greater evidence of adequate measurement properties. Except for the Healthy Pathways and CHU-9D, these PROMs have been tested with children and young people affected by neurodisability; the evidence shows a similar pattern albeit supported by fewer studies. Most noticeably absent for all these PROMs are studies examining content validity. Thus these PROMs might be leading candidates for further testing in groups with neurodisability, particularly the properties of responsiveness and longitudinal validity. Tests of responsiveness and longitudinal validity assess how scale scores change over time and whether the direction and magnitude of the changes reflect what would be expected on the basis of theory determined in advance, ideally incorporating a comparison with a group not expected to change. In the absence of evidence of responsiveness, those selecting PROMs should appraise whether the aspects of health assessed by tools and the response options to questions suggest that these are ‘likely’ to change in their specific context for application.

Although PROMs are generally designed for use as group measures in service evaluations, audits, and research, there is also growing interest in using them clinically as individualized measures. The proposed criterion for test–retest reliability is more stringent for individualized use (intraclass correlation coefficient >0.9), and such high levels of stability would need to be demonstrated in paediatric neurodisability.

Aside from the standard measurement properties, there are several other criteria that apply when selecting a candidate PROM. These include appropriateness, acceptability to potential respondents, and feasibility: for example the burden on respondents and those administering and processing data. We studied the appropriateness of existing generic and chronic-generic PROMs for children with neurodisability by asking whether they cover the more important aspects of health for this particular group. We sought to identify a core set of outcomes that could be assessed using PROMs for these children; that is, outcomes beyond mortality and morbidity. To this end we performed qualitative research separately with children and parents, a Delphi survey with health professionals, and held a prioritization meeting. This work produced a core set of outcomes deemed important to children and/or parents that were aspects of health targeted by National Health Service clinicians. The domains were communication, emotional wellbeing, pain, sleep, mobility, self-care, independence, mental health, community and social life, behaviour, toileting, and safety. However, none of the identified PROMs capture all these key domains. Adding to this the scarce evidence of good overall psychometric performance for existing measures in a population with neurodisability, there could be a place to refine or develop existing PROMs accordingly.

There are some limitations to this systematic review; most are a consequence of the strict inclusion criteria. Neurodisability comprises a vast number of conditions, and although we included other general descriptions and MeSH (Medical Subject Headings) terms for developmental disabilities, we only had three key marker conditions (CP, autism, and epilepsy) and relevant variations on neuro-motor, neuropsychiatric, and developmental disabilities. Although we updated searches for evaluation studies up to 30 July 2014, we did not repeat the systematic search to identify any new PROMs after September 2012. Hence, we will not have included any new PROMs published after this date; however, we are not aware of any such PROMs that would meet the eligibility criteria.

One of our inclusion criteria was published peer-reviewed reports of studies that specifically set out to evaluate measurement properties of PROMs. Hence, we excluded papers that might have presented incidental evidence from studies where PROMs were used in observational or experimental studies. However, information from studies that were not designed specifically to test measurement properties can be misleading. Studies testing responsiveness require testing of some a priori hypothesis in a longitudinal study, whereas evaluative trials typically test interventions of unknown effectiveness. Therefore, for instance, observing no change could be interpreted as either a blunt, non-responsive measure or an ineffective intervention, and it is not possible to determine which is true. In addition, we will have omitted any information that may be contained in manuals, if these data have not been published in peer-reviewed journals. We justify this as peer review provides some level of quality assurance to the evidence being appraised. We included studies with children and young people with chronic conditions, providing the samples included neurodisability. Hence, we did not appraise studies examining PROMs with children with other conditions (e.g. arthritis or asthma).
Limiting the review to studies where an English version of the PROM was administered excluded some PROMs from further analyses. Two PROMs excluded from this review that may warrant further investigation are ITQoL (for infants), which was developed in the Netherlands and for which an English translation is available but no published studies of this version were found, and the TNO-AZL (TACQOL, TAPQOL, and TAAQOL). If studies had been included that used versions of questionnaires in languages other than English, then further evidence would have emerged, for instance regarding the KINDL and the plethora of translated versions of the more popular instruments such as PedsQL. Nevertheless, psychometric performance cannot be assumed across languages and cultures; therefore, in our view, limiting the review to evaluations of English-language versions is a relative strength of it.

There remains much scope for research in evaluating multidimensional PROMs to measure health outcomes in paediatric neurodisability, particularly in testing item invariance across conditions and the responsiveness of PROM scores to quantify meaningful change that is beyond measurement error.

ACKNOWLEDGEMENTS
This study was part of research funded by the NIHR Health Services and Delivery Research programme (project 10/0202/16 http://www.nets.nihr.ac.uk/projects/hsdr/10200216). The work also benefited from support from NIHR PenCLAHRC and the Charity Cerebra. The views and opinions expressed in this paper are those of the authors and not necessarily those of the National Health Service, the NIHR, the Department of Health, or Cerebra. We are grateful to co-investigators in the broader project, particularly Colin Green and Jo Thompson Coon for their contribution to the development of the protocol, and Anna Stimson for her administrative support. The authors have stated that they had no interests that could be perceived as posing a conflict or bias.

SUPPORTING INFORMATION
The following additional material may be found online:
Table S1: PROMs (group of questionnaires), the different versions (according to age group, length, or responder), acronyms, and reference citations, including reference citation.
Appendix S1: An example of the search strategy used on Ovid MEDLINE(R), In-Process & Other Non-Cited Indexes, and Ovid MEDLINE(R) (1946 to present).

REFERENCES
1. US Food and Drug Administration. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville, MD: Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, 2009.
2. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. http://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf (accessed 23 September 2014).
3. Fitzpatrick R. In: Smith PC, Mossialos E, Papanicolas I, Carlon S, Shields N, Yong K, Gilmore R, Sakzewski E, editors. Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects. Cambridge University Press, 2009: 63–86.
4. Morris C, Janssens A, Tomlinson R, Williams J, Logan S. Towards a definition of neurodisability: a Delphi survey. Dev Med Child Neurol 2011; 55: 1103–08.
5. Carlson S, Shields N, Yong K, Gilmore R, Sakzewski L, Boyd R. A systematic review of the psychometric properties of quality of life measures for school aged children with cerebral palsy. BMC Pediatr 2010; 10: 81.
6. Sadeghi S, Fayed N, Ronen GM. Patient-reported outcome measures in pediatric epilepsy: a content analysis using World Health Organization definitions. Epilepsia 2014; 55: 1411–17.
7. Rosenbaum P. The ABCs of clinical measures. Dev Med Child Neurol 2015; 57: 496.
8. Fayed N, de Camargo OK, Kerr E, et al. Generic patient-reported outcomes in child health research: a review of conceptual content using World Health Organization definitions. Dev Med Child Neurol 2012; 54: 1085–95.
9. Janssens A, Thompson Coon J, Rogers M, et al. A systematic review of generic multidimensional patient-reported outcome measures for children, part I: descriptive characteristics. Value Health 2015; 18: 315–33.
10. Wild D, Eremenco S, Mear I, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. Value Health 2009; 12: 430–40.
11. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force on Translation and Cultural Adaptation. Value Health 2015; 8: 94–104.
12. Strenner DL, Norman GR. Health Measurement Scales: A Practical Guide to their Development and Use. 3rd edn. Oxford: Oxford University Press, 2003.
13. Mokkink LB, Terwee CB, Knol DL, et al. The COS-MIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010; 10: 22.
14. Mokkink LB, Terwee CB, Patrick DL, et al. The COS-MIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63: 737–45.
15. Janssens A, Rogers M, Thompson Coon J, et al. A systematic review of generic multidimensional patient-reported outcome measures for children, part II: evaluation of psychometric performance of English-language versions in a general population. Value Health 2015; 18: 334–45.
16. Mokkink LB, Terwee CB, Patrick DL, et al. The COS-MIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010; 19: 519–49.
17. Fitzpatrick R, Davey C, Butson MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998; 2: i–iv, 1–74.
18. Uijen AA, Heinot CW, Schellevis FG, et al. Measurement properties of questionnaires measuring continuity of care: a systematic review. PLoS ONE 2012; 7: e42256.
19. Davies N, Mackintosh A, Gibbons E, Fitzpatrick R. A structured review of patient-reported outcome measures for women with breast cancer. (Appendix Bii). Patient-reported Outcome Measurement Group, University of Oxford, Oxford, UK, 2009. http://phlt.uoeh.ac.uk/ (accessed 10 February 2015).
20. McDowell I, Newell C. Measuring Health. A Guide to Rating Scales and Questionnaires. New York, NY: Oxford University Press, 1996.
21. Yarni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life?: an analysis of 8,591 children across age subgroups with the PedsQL-xt 4.0 Generic Core Scales. Health Qual Life Outcomes 2007; 5: 1.
22. Yarni JW, Burwinkle TM. The PedsQL as a patient-reported outcome in children and adolescents with attention-deficit/hyperactivity disorder: a population-based study. Health Qual Life Outcomes 2006; 4: 26.
23. Davis SE, Hynan LS, Limbers CA, et al. The PedsQL in pediatric patients with Duchenne muscular dystrophy: feasibility, reliability, and validity of the Pediatric Quality of Life Inventory Neuromuscular Module and Generic Core Scales. J Clin Neurosci 2010; 11: 97–109.

24. Limbers CA, Ripperger-Suhler J, Heffer RW, Varni JW. Patient-reported pediatric quality of life inventory 4.0: generic core scales in pediatric patients with attention-deficit/hyperactivity disorder and comorbid psychiatric disorders: feasibility, reliability, and validity. Value Health 2011; 14: 521–30.

25. Davis E, Shelly A, Waters E, Davern M. Measuring the quality of life of children with cerebral palsy: comparing the conceptual differences and psychometric properties of three instruments. Dev Med Child Neurol 2010; 52: 174–80.

26. Iannaccone ST, Hynan LS, Morton A, et al. 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: 805–12.

27. Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children’s health-related quality of life: an analysis of 13,878 parents’ reliability and validity across age subgroups using the PedsQL® Jr 4.0 Generic Core Scales. Health Qual Life Outcomes 2007; 5: 2.

28. Oeltzing O, Bagley A, Rogers S, et al. Outcome tools used for ambulatory children with cerebral palsy: responsiveness and minimum clinically important differences. Dev Med Child Neurol 2008; 50: 918–25.

29. Rentz AM, Matz LS, Secnik K, Swensen A, Revicki DA. Psychometric validation of the child health questionnaire (CHQ) in a sample of children and adolescents with attention-deficit/hyperactivity disorder. Qual Life Res 2005; 14: 719–34.

30. Wake M, Salomon L, Reddihough D. Health status of Australian children with mild to severe cerebral palsy: cross-sectional survey using the Child Health Questionnaire. Dev Med Child Neurol 2003; 45: 194–99.

31. McCullough N, Parkes J, White-Koning M, Beckung E, Colver A. Reliability and validity of the Child Health Questionnaire PF-50 for European children with cerebral palsy. J Pediatr Psychol 2009; 34: 41–50.

32. Chaplin JE, Koopman HM, Schmidt S. DISABKIDS Smiley Questionnaire: the TAKE 6 assisted health-related quality of life measure for 4 to 7-year-olds. Clin Psychol Psychother 2008; 15: 173–80.

33. Erhart M, Ravens-Sieberer U, Dickinson HO, Cober A. European SPARCLE and Kidscreen Groups. Rasch measurement properties of the KDSCSCREEN quality of life instrument in children with cerebral palsy and differential item functioning between children with and without cerebral palsy. Value Health 2009, 12: 782–92.

34. McDougall J, Wright V, Nicholls M, Miller L. Assessing the psychometric properties of both a global and a domain-specific perceived quality of life measure when used with youth who have chronic conditions. Soc Indic Res 2013; 114: 1249–57.

35. Riley AW, Coghill D, Forrest CB, Lorenzo MJ, Ralston SJ, Spiel G. Validity of the health-related quality of life assessment in the ADORE study: parent report form of the CHIP-Child Edition. Eur Child Adolesc Psychiatry 2006; 15(Suppl. 1): 63–71.

36. Schacht A, Escoar R, Wagner T, Wehmeier PM. Psychometric properties of the quality of life scale Child Health and Illness Profile-Child Edition in a combined analysis of five amoebomine trials. Atti Defis Hyperact Disabil 2011; 3: 335–49.

37. Edwards TC, Huebner CE, Connell FA, Patrick DL. Adolescent quality of life, part I: conceptual and measurement model. J Adolesc 2002; 25: 275–86.

38. Lai JS, NowinsKi C, Victorson D, et al. Quality-of-life measures in children with neurological conditions: pedi- atric Neuro-QOL. Neurorehabil Neural Repair 2012; 26: 36–47.

39. Cella D, Nowinski C, Peterman A, et al. The neurology quality-of-life measurement initiative. Arch Phys Med Rehabil 2011; 92: S28–36.

40. Perez L, Huang J, Jansky L, et al. Using focus groups to inform the Neuro-QOL measurement tool: exploring patient-centered, health-related quality of life concepts across neurological conditions. J Neurosci Nurs 2009; 39: 342–51.

41. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertainning the minimal clinically important difference. Control Clin Trials 1989; 10: 407–15.

42. Hillard ME, Lawrence JM, Modi AC, et al. Identification of minimal clinically important difference scores of the PedsQL in children, adolescents, and young adults with type 1 and type 2 diabetes. Diabetics Care 2013; 36: 1891–97.

43. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. Phys Ther 2006; 86: 735–43.

44. Cohen J. Statistical Power Analysis for the Behavioural Sciences. Mahwah, NJ: Lawrence Erlbaum Associates, 1988.

45. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal impor- tant differences for patient reported outcomes. Health Qual Life Outcomes 2006; 4: 70.

46. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. Health Econ 1999; 8: 41–51.

47. Drummond MF, Sculpher MJ, Torrance GW, O’Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. 3rd edn. Oxford: Oxford University Press, 2005.

48. Bevans KB, Riley AW, Forrest CB. Development of the healthy pathways child-report scales. Qual Life Res 2010; 19: 1195–214.

49. Stevens KJ. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. Qual Life Res 2009; 18: 1105–13.

50. Santana MJ, Haverman L, Absolom K, et al. Training clinicians in how to use patient-reported outcome mea- sures in routine clinical practice. Qual Life Res 2015; 24: 1707–18.

51. Nelson EC, Efmimova E, Lind C, Hager A, Wason JHL, Lindblad S. Patient reported outcome measures in practice. BMJ 2015; 350: g7818.

52. Allard A, Fellowes A, Shilling V, Janssens A, Beresford B, Morris C. Key health outcomes for children and young people with neurodisability: qualitative research with young people and parents. BMJ Open 2014; 4: e004611.

53. Janssens A, Williams J, Tomlinson R, Logan S, Morris C. Health outcomes for children with neurodisability: what do professionals regard as primary targets? Arch Dis Child 2014; 99: 927–32.

54. Morris C, Janssens A, Shilling V, et al. Meaningful health outcomes for paediatric neurodisability: Stake- holder prioritisation and appropriateness of patient reported outcome measures. Health Qual Life Outcomes 2015; 13: 87.

55. De Vet HC, Terwee C, Mokkink LB, Knol DL. Mea- surement in Medicine: A Practical Guide. Cambridge: Cambridge University Press, 2011.

56. Klassen AF, Landgraf JM, Lee SK, et al. Health related quality of life in 3 and 4 year old children and their par- ents: preliminary findings about a new questionnaire. Health Qual Life Outcomes 2003; 1: 81.

57. Fekkes M, Theunissen NCM, Brugman E, et al. De- velopment and psychometric evaluation of the TAPOQOL: a health-related quality of life instrument for 1–5-year-old children. Qual Life Res 2000; 9: 961–72.

58. Ravens-Sieberer U, Ballinger M. Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analyti- cal results. Qual Life Res 1998; 7: 399–407.

59. Landgraf JM, Abetz LN. Functional status and well- being of children representing three cultural groups: initial self-reports using the CHQ-CF87. Psychol Health 1997; 12: 839–54.

60. Pencharz J, Young NL, Owen JL, Wright JG. Compari- son of three outcome instruments in children. J Pediatr Orthop 2001; 21: 425–32.

61. Vitale MG, Levy DE, Moskowitz AJ, et al. Capturing quality of life in pediatric orthopaedics: two recent mea- sures compared. J Pediatr Orthop 2001; 21: 629–35.

62. Vitale MG, Roye EA, Choe JC, Hyman JE, Lee FY, Roye DP Jr. Assessment of health status in patients with cerebral palsy: what is the role of quality-of-life mea- sures? J Pediatr Orthop 2005; 25: 792–97.

63. Thomas-Stonell N, Johnson P, Runney P, Wright V, Oddson B. An evaluation of the responsiveness of a comprehensive set of outcome measures for children and adolescents with traumatic brain injuries. Pediatr Rehabil 2006; 9: 14–23.

64. Drotar D, Schwartz L, Palermo TM, Burant C. Factor structure of the Child Health Questionnaire-Parent Form in pediatric populations. J Pediatr Psychol 2006; 31: 127–38.

65. Ferro MA, Landgraf JM, Speckley KN. Factor struc- ture of the Child Health Questionnaire Parent Form-50 and predictors of health-related quality of life in chil- dren with epilepsy. Qual Life Res 2013; 22: 2201–02.

66. Saigal S, Rosenbaum P, Stokopf B, et al. Development, reliability and validity of a new measure of overall health for pre-school children. Qual Life Res 2005; 14: 243–57.

67. Graham P, Stevenson J, Flynn D. A new measure of health-related quality of life for children: preliminary findings. Psychol Health 1997; 12: 655–65.
68. Petersen C, Schmidt S, Power M, Bullinger M, Group D. Development and pilot-testing of a health-related quality of life chronic generic module for children and adolescents with chronic health conditions: a European perspective. Qual Life Res 2005; 14: 1065–77.
69. Schmidt S, Debensason D, Muhlan H, et al. The DISABKIDS generic quality of life instrument showed cross-cultural validity. J Clin Epidemiol 2006; 59: 587–98.
70. Simeoni M-C, Schmidt S, Muehlan H, Debensason D, Bullinger M, Group D. Field testing of a European quality of life instrument for children and adolescents with chronic conditions: the 37-item DISABKIDS chronic generic module. Qual Life Res 2007; 16: 881–93.
71. Matza LS, Secnik K, Mannix S, Sallee FR. Parent-proxy EQ-5D ratings of children with attention-deficit hyperactivity disorder in the US and the UK. Pharmacoeconomics 2005; 23: 777–90.
72. Glaser AW, Furlong W, Walker DA, et al. Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the U.K. Eur J Cancer 1999; 35: 256–61.
73. Tilford JM, Payakachat N, Kovacs E, et al. Preference-based health-related quality-of-life outcomes in children with autism spectrum disorders: a comparison of generic instruments. Pharmacoeconomics 2012; 30: 661–79.
74. Eser C, Vance YH, Horne B, Glaser A, Galvin H. The value of the Pediatric Quality of Life in assessing quality of life in survivors of childhood cancer. Child Care Health Dev 2003; 29: 95–102.
75. McCarthy ML, MacKenzie IJ, Durbin DR, et al. The Pediatric Quality of Life Inventory: an evaluation of its reliability and validity for children with traumatic brain injury. Arch Phys Med Rehabil 2005; 86: 1901–49.
76. Varni JW, Burwinkle TM, Berrin SJ, et al. The PedsQL in pediatric cerebral palsy: reliability, validity, and sensitivity of the Generic Core Scales and Cerebral Palsy Module. Dev Med Child Neurol 2006; 48: 442–49.
77. Palmer SN, Meeske KA, Katz ER, Burwinkle TM, Varni JW. The PedsQL Brain Tumor Module: initial reliability and validity. Pediatr Blood Cancer 2007; 49: 287–93.
78. Majnemer A, Shovel M, Law M, Poulin C, Rosenbaum P. Reliability in the ratings of quality of life between parents and their children of school age with cerebral palsy. Qual Life Res 2008; 17: 1163–71.
79. Varni JW, Limbers CA, Newman DA. Using factor analysis to confirm the validity of children’s self-reported health-related quality of life across different modes of administration. Clin Transl 2009; 6: 185–95.
80. Young NL, Varni JW, Snider L, et al. The Internet is valid and reliable for child-report: an example using the Activities Scale for Kids (ASK) and the Pediatric Quality of Life Inventory (PedsQL). J Clin Epidemiol 2009; 62: 314–20.
81. Limbers CA, Heffer RW, Varni JW. Health-related quality of life and cognitive functioning from the perspective of parents of school-aged children with Autism spectrum utilizing the PedsQL. J Autism Dev Disord 2009; 39: 1529–41.
82. Dunaway S, Montes J, Montgomery M, et al. Reliability of telephone administration of the PedsQL Generic Quality of Life Inventory and Neuromuscular Module in spinal muscular atrophy (SMA). Neuromuscul Disord 2010; 20: 162–65.
83. Shipman DL, Sheldrick RC, Perrin EC. Quality of life in adolescents with autism spectrum disorders: reliability and validity of self-reports. J Dev Behav Pediatr 2011; 32: 85–89.
84. Green L, Godfrey C, Soo C, Anderson V, Cattozza C. Agreement between parent-adolescent ratings on psychosocial outcome and quality-of-life following childhood traumatic brain injury. Dev Neuropsychol 2012; 15: 105–13.
85. Sheldrick R, Neger EN, Shipman D, Perrin EC. Quality of life of adolescents with autism spectrum disorders: concordance among adolescents’ self-reports, parents’ reports, and parents’ proxy reports. Qual Life Res 2012; 21: 53–57.
86. Tavormar L, Barron E, Rodgers J, McNamara H. Finding out what matters: validity of quality of life measurement in young people with ASD. Child Care Health Dev 2013; 39: 592–601.
87. Patrick DL, Edwards TC, Topolski TD. Adolescent quality of life, part II: initial validation of a new instrument. J Adolesc 2002; 25: 287–300.
88. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorder. 4th edn. Washington, DC: American Psychiatric Association, 2000.