

REVIEW

Open Access



A critical evaluation of the content validity of patient-reported outcome measures assessing health-related quality of life in children with cancer: a systematic review

Maria Rothmund^{1,2*} , Andreas Meryk³, Gerhard Rumpold¹, Roman Crazzolaro³, Samantha Sodergren⁴, Anne-Sophie Darlington⁴ and David Riedl^{1,5} on behalf of the EORTC Quality of Life Group

Abstract

Background With increasing survival rates in pediatric oncology, the need to monitor health-related quality of life (HRQOL) is becoming even more important. However, available patient-reported outcome measures (PROMs) have been criticized. This review aims to systematically evaluate the content validity of PROMs for HRQOL in children with cancer.

Methods In December 2021, a systematic literature search was conducted in PubMed. PROMs were included if they were used to assess HRQOL in children with cancer and had a lower age-limit between 8 and 12 years and an upper age-limit below 21 years. The COSMIN methodology for assessing the content validity of PROMs was applied to grade evidence for relevance, comprehensiveness, and comprehensibility based on quality ratings of development studies (i.e., studies related to concept elicitation and cognitive interviews for newly developed questionnaires) and content validity studies (i.e., qualitative studies in new samples to evaluate the content validity of existing questionnaires).

Results Twelve PROMs were included. Due to insufficient patient involvement and/or poor reporting, the quality of most development studies was rated 'doubtful' or 'inadequate'. Few content validity studies were available, and these were mostly 'inadequate'. Following the COSMIN methodology, evidence for content validity was 'low' or 'very low' for almost all PROMs. Only the PROMIS Pediatric Profile had 'moderate' evidence. In general, the results indicated that the PROMs covered relevant issues, while results for comprehensiveness and comprehensibility were partly inconsistent or insufficient.

Discussion Following the COSMIN methodology, there is scarce evidence for the content validity of available PROMs for HRQOL in children with cancer. Most instruments were developed before the publication of milestone guidelines and therefore were not able to fulfill all requirements. Efforts are needed to catch up with methodological progress made during the last decade. Further research should adhere to recent guidelines to develop new instruments and to strengthen the evidence for existing PROMs.

Keywords Patient-reported outcomes, Content validity, Questionnaire development, Pediatric oncology, Childhood cancer, Psychometrics, Patient-reported outcome measures

*Correspondence:

Maria Rothmund

Maria-Sophie.Rothmund@i-med.ac.at

Full list of author information is available at the end of the article



© The Author(s) 2023. **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/>.

Background

In recent decades, survival rates in pediatric oncology have increased considerably [1–3]. Even though overall survival remains the primary outcome [4], patients' health-related quality of life (HRQOL) also needs careful monitoring and management. HRQOL as defined by the World Health Organization (WHO) is an "individual's perception of their position in life [...] incorporating in a complex way individuals' physical health, psychological state, level of independence, social relationships, personal beliefs and their relationships to salient features" [5]. Depending on context and target population, different aspects are relevant for HRQOL. For children with cancer, Anthony et al. [6] have provided the most comprehensive conceptual framework so far. It covers four major domains: physical (symptoms, physical functioning), psychological (emotional distress, behavior, positive psychological function, self-esteem, body image, cognitive health), social (relationships, social functioning), and general health (health perception) [6].

In clinical routine and research, HRQOL is commonly assessed by patient-reported outcome measures (PROMs). In pediatrics, PROMs are often complemented with caregiver-reports. However, patient- and caregiver-reports often differ, especially for less observable outcomes that are only accessible from patient perspective (e.g., perceived burden, satisfaction with relationships) [7–12]. Several studies have indicated that children from 8 years onwards can reliably self-report [13–15]. Thus, it is recommended to treat patient-reports as the most important source of information in this age-group [7, 16]. This is in line with a trend towards increasing the involvement and empowerment of children in research and treatment [17–19].

To assess HRQOL from children's perspective, evidence-based and age-appropriate PROMs are needed that meet psychometric quality criteria [20]. The most fundamental measurement property is content validity, defined as "the degree to which the content [...] is an adequate reflection of the construct(s) to be measured" [20]. Claims regarding content validity can only be made when an instrument *comprehensively* assesses *relevant* aspects in a *comprehensible* way [21, 22].

To ensure content validity, PROM development guidelines strongly recommend patient involvement in several stages [15, 21, 23–26]. They suggest involving patients in concept elicitation and issue generation to give their opinion on relevance and comprehensiveness. Later in the process, guidelines request cognitive interviews to evaluate whether item formulations, response-options, and recall-periods are understood as intended.

For children from the age of 8 years, recall-periods from 7 days to 4 weeks and faces-scales with ≤ 6 faces or

Likert-scales with ≤ 5 points are usually considered suitable [24, 27]. Adolescents and young adults (AYAs) around 14 years or older can complete the same tools as adults [28], but they face distinct HRQOL issues as they transition into adulthood [29, 30].

Previous research has indicated that children with cancer were insufficiently involved in the development of existing PROMs [31]. It has been questioned whether they measure what is relevant for children [32], and whether they are complete [33] and of sufficient psychometric quality [31, 34].

The present systematic review aims to systematically evaluate the content validity of available PROMs for HRQOL in children with cancer aged between 8 and 14 years. To do so, the COSMIN methodology for assessing the content validity of PROMs [21, 22; COSMIN = COnsensus-based Standards for the selection of health Measurement INstruments] is applied. In a recently published review, this methodology was used to evaluate PROMs measuring positive psychological constructs [35]. Previous reviews using the COSMIN methodology to evaluate PROMs for pediatric oncology [34, 36, 37] were based on an older version [38–40], which was less comprehensive. The previous COSMIN guideline did not cover the key concept of comprehensibility, and its standards only checked whether certain steps were undertaken, without evaluating the methodological quality [22]. Thus, it is expected that ratings based on the old version will vary considerably from ratings based on the current version.

Methods

This systematic review follows the Preferred Reporting Items of Systematic Reviews and Meta-analyses (PRISMA) guidelines, where applicable [41]. The PRISMA checklist is provided in Additional file 1. At the time when we started to work on this review, it was not possible to register the protocol since common platforms (e.g., PROSPERO) accepted COVID-19-related protocols only. Thus, no protocol has been published.

Search strategy and study selection

A literature search was conducted on PubMed in December 2021 combining Medical Subject Headings (MeSH) related to HRQOL, the target population of children with cancer, and psychometrics: ("Quality of Life"[MeSH] AND (Neoplasms [MeSH] OR "Medical Oncology"[MeSH]) AND (Child [MeSH] OR Pediatrics [MeSH]) AND ("Self Assessment"[MeSH] OR "Patient Reported Outcome Measures"[MeSH] OR "Patient Outcome Assessment"[MeSH] OR "Self Report"[MeSH] OR "Psychometrics"[MeSH])). The search was neither

limited to a specific time-period nor filtered for specific languages.

As a first step, abstracts were screened by one reviewer [MR] to identify PROMs for HRQOL assessment used in children with cancer within the age range between 8 and 14 years. This included generic and cancer-specific instruments but excluded survivor-specific instruments. PROMs primarily addressing adolescents (lower age-limit at ≥ 12) were excluded, but PROMs for transitional age-groups (children and adolescents) were included if the upper age-limit did not exceed 21 years. A PROM was considered relevant if the developers claimed to assess HRQOL or if it covered physical, psychological, and social health, as described in the conceptual framework by Anthony et al. [6]. PROMs assessing single symptoms or adverse effects were excluded (e.g., PedsQL Fatigue scale [42] or separate PROMIS-scales [43]).

To ensure that all relevant PROMs were included, the list of PROMs was compared to a list of 112 instruments identified by Algu ren et al. for the development of the Overall Pediatric Health Standard Set (OPH-SS) [44] and a list of 155 PROMs collected in a simultaneously conducted review of HRQOL issues in children with cancer [45]. For all included instruments, manuals and review copies were searched. If not accessible, authors were contacted. Data regarding their main characteristics were extracted [MR], i.e., the target population (age, diagnoses), recall-period, response-options, the number of items, and the intended scale structure as well as whether a parent-version was available (see Table 1).

In a second step, full-texts and their reference-lists were screened by one reviewer [MR] to identify development and content validity studies for the investigated PROMs. The inclusion and exclusion criteria were based on the definitions provided by the COSMIN guidelines: Development studies include all studies on concept elicitation and studies testing PROMs under development, e.g., cognitive interview studies. Content validity studies include all studies that investigate the relevance, comprehensiveness, and/or comprehensibility of existing PROMs in a new sample. Additional searches on PubMed were conducted with PROM-names and “*develop**” or “*content valid**” to check whether further relevant studies were available. The included studies were evaluated according to the COSMIN guidelines (see below).

The COSMIN methodology for assessing content validity

The COSMIN methodology for assessing content validity is divided into three so-called ‘boxes’ with several ‘standards’ [22, 46]. Box 1 evaluates the quality of PROM development, including general design (definition of construct, target population, and context/purpose; 35

standards), concept elicitation (7 standards), and cognitive interviews (22 standards).

Box 2 evaluates the quality of content validity studies, defined as studies on the relevance, comprehensiveness, and comprehensibility of existing PROMs performed in new samples [22]. The standards in box 2 assess whether and how patients were asked about relevance (standards 1–7), comprehensiveness (standards 8–14), and comprehensibility (standards 15–21), and whether and how professionals were asked about relevance (standards 22–26) and comprehensiveness (standards 27–31). As caregivers play an important intermediary role in pediatrics, we wanted to take their input into account as well. After consulting with the COSMIN Group, we decided to use the standards for expert involvement (standards 22–31) to rate whether and how caregivers were asked about relevance and comprehensiveness.

In box 3, the results of development and content validity studies are rated against ten criteria for good content validity. Additionally, reviewers were asked to give their own ratings of comprehensiveness, relevance, and comprehensibility of the tool (eight standards). In terms of comprehensibility, ratings for response-options and recall-periods were based on recommendations from a recent review by Coombes et al. [27]. Item-formulations were rated positive, except if items appeared obviously inappropriate for children. For consistent relevance and comprehensiveness ratings, the items of all PROMs were systematically categorized by content, as described below.

In a final step, the overall ratings are summarized and the quality of evidence is graded. Following the COSMIN guidelines, evidence is rated ‘low’ or ‘very low’ if there has been no content validity study of at least ‘doubtful’ quality. If content validity has not been sufficiently assessed, the development process needs to be of ‘adequate’ or ‘very good’ quality to obtain a ‘moderate’ evidence level. For evidence to obtain a ‘high’ rating, there needs to have been at least one content validity study of ‘adequate’ or ‘very good’ quality.

The ratings of boxes 1 and 2 were conducted by two reviewers independently [MR, AM], using the Excel-sheet available from the COSMIN website (cosmin.nl). We made minor adaptations to this sheet by adding columns for the reviewers to justify their decisions. Conflicts were discussed until consensus was reached. The ratings of box 3 and the final evidence grading were performed by one reviewer [MR] and approved by all co-authors.

Categorizing items by the contents assessed

To provide a uniform and solid basis for reviewers’ ratings of comprehensiveness and relevance, items

Table 1 Main characteristics of the included Patient-Reported Outcome Measures (PROMs)

PROM	Target population	Parent-version	Recall period	Response options	Intended scale structure Scale name (number of items per scale) ^a	Studies taken into account
DISABKIDS 12/37	8–16 Chronic Disease	Yes	4 weeks	5-point Likert-scale	Physical limitation (2/6) Treatment (2/6) Independence (2/6) Emotions (2/7) Social exclusion (2/6) Social inclusion (2/6)	[52–55, 68–71]
KIDSCREEN 10/27/52	8–18 Generic	Yes	Last week	5-point Likert-scale	KIDSCREEN 10: Unidimensional trait (10) Overall health perception (1) KIDSCREEN 27: Physical well-being (5) Psychological well-being (7) Autonomy and parents (7) Social support and peers (4) School environment (4) KIDSCREEN 52: Physical well-being (5) Psychological well-being (6) Moods and emotions (7) Self-perception (5) Autonomy (5) Parent relations and home (6) Social support and peers (6) School environment (6) Social acceptance/ bullying (3) Financial resources (3)	[47, 48, 72–74]
KINDL-R Generic Kid	7–13 Generic	Yes	Past week	5-point Likert-scale	Physical (4) Emotional (4) Self-esteem (4) Family (4) Friends (4) School (4)	[49, 50, 75–77]
KINDL-R Oncology Module Kid	7–17 Cancer	No	Past week	3-/4-/5-point Likert-scale	Physical well-being (4) Psychological well-being (4) Friends (3) School (2) Treatment (11)	[60]
PAC-QoL Child (provisional)	8–12 Advanced cancer	No	Last week	4-point Likert-scale	Physical comfort (?) Psychological well-being (?) Social interaction (?) Resilience (?) Quality of care (?)	[61, 62]

Table 1 (continued)

PROM	Target population	Parent-version	Recall period	Response options	Intended scale structure Scale name (number of items per scale) ^a	Studies taken into account
PedsQL Brain Tumor Module Child	8–12 Brain Tumor	Yes	7 days	5-point Likert-scale	Cognitive problems (7) Pain and hurt (3) Movement and balance (3) Procedural anxiety (3) Nausea (5) Worry (3)	[63, 78–81]
PedsQL Cancer Module 3.0 Child	8–12 Cancer	Yes	one month	5-point Likert-scale	Pain and hurt (2) Nausea (5) Procedural anxiety (3) Treatment anxiety (3) Worry (3) Cognitive problems (5) Perceived physical appearance (3) Communication (3)	[42, 78–80]
PedsQL Generic Core Scale 4.0 Child	8–12 Generic	Yes	one month	5-point Likert-scale	Health and activities (8) Feelings (5) Social functioning (5) School (5)	[42, 51, 78–80, 82–84]
PROMIS v2.0 Pediatric Profile 25/37/49	8–17 Generic and/or chronic Disease	Yes	7 days	5-point Likert-scale	Physical function mobility (4/6/8) Anxiety (4/6/8) Depressive symptoms (4/6/8) Fatigue (4/6/8) Peer relationships (4/6/8) Pain interference (4/6/8) Pain intensity (1)	[56, 57, 85–99]
QOLCC-7-12	7–12 Cancer	Yes	N/A	N/A	Physical function (5) Psychological function (6) Social function (7) Treatment/symptoms (6) Cognitive function (6) Understanding illness (4) Patient-communication (6)	[64, 65, 100, 101]
SQOLPOP	7–12 Cancer	Yes	N/A	N/A	N/A	[66, 67]
TACQOL	6–15 Chronic Disease	Yes	2 weeks	4-point Likert-scale	Body (8) Motor (8) Autonomy (8) Cognition (8) Social (8) Positive emotions (8) Negative emotions (8)	[58, 59, 102]

(?) not unambiguously clear which—and therefore how many—items belong to which (sub)scale

N/A: No information available

^a Numbers separated by slash: numbers of items per scale for different length versions; sorted from the shortest version to the longest version

from all investigated PROMs were extracted into an Excel-file and mapped onto the conceptual framework by Anthony et al. [6]. Within this hierarchical framework, the domains of physical, psychological, and social health were further divided into subdomains, containing several identifying concepts. For example, physical health is divided into symptoms (e.g., pain, fatigue) and physical function (e.g., dexterity, mobility), while social health is divided into relationships (e.g., with family or peers) and social function (e.g., recreation and leisure, school). The psychological domain has the most subdomains and is divided into emotional distress (e.g., afraid, sad), behavior (e.g., clingy, defiant), positive psychological function (e.g., benefit finding), self-esteem (e.g., feeling loved or proud), body image (e.g., personal appearance), and cognitive issues (e.g., attention, remembering).

Each item was assigned to one domain, subdomain, and identifying concept by one reviewer [MR]. Open-ended questions, conditional items (filter-questions), and determinant questions (on background information of the patient) were not taken into account. To enable a consistent categorization across all items, we defined categorization rules (Additional file 2). A second reviewer [DR] indicated his (dis)agreement per item. Conflicts were discussed until consensus was reached. Where necessary, new subdomains and identifying concepts were

added to complement the conceptual framework (Additional file 3).

Descriptive statistics were applied to investigate the representation of contents within the overall item pool and the questionnaires. Item content was considered relevant if it could be assigned to one of the subdomains. Questionnaires were considered comprehensive when they covered physical health and social health (at least family/general) and several aspects of psychological health, i.e., negative emotional health issues (emotional distress or treatment burden), positive issues (positive psychological functioning or self-esteem), and cognitive issues.

Results

Identification of PROMs and their main characteristics

As shown in Fig. 1, the literature search identified 231 articles and screening for PROMs resulted in a list of nine inventories (i.e. measurement systems / questionnaire providers). Two of them provided different modules (e.g., generic and cancer-specific), resulting in 12 different PROMs. Taking versions of different length into account, 17 questionnaires were identified. Counterchecking against the PROMs collected for the development of the OPH-SS [44] and our review of HRQOL issues [45] did not yield any additional instruments. For the included PROMs, 53 development and content validity studies

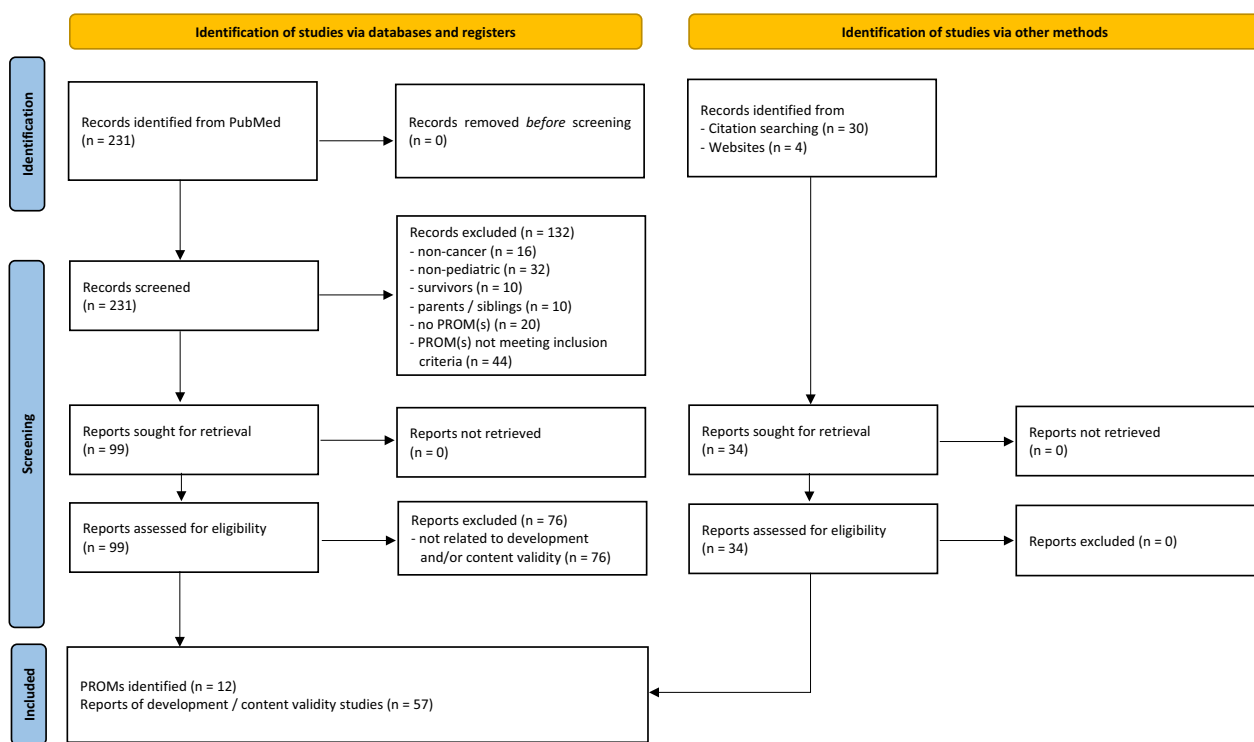


Fig. 1 PRISMA 2020 flow diagram of articles selection process. From: Page et al. [41]

and four manuals were identified that were taken into account in the present evaluation (Table 1).

Among the 12 PROMs, three are generic instruments (KIDSCREEN [47, 48], KINDL-R Kid Generic [49, 50], PedsQL Generic Core Scale [42, 51]), another three are for chronically ill children (DISABKIDS [52–55], PROMIS Pediatric Profile [56, 57], and TACQOL-CF [58, 59]), and six are cancer-specific (KINDL-R Kid Oncology Module [60], PAC-QoL Child [61, 62], PedsQL Brain Tumor [63], PedsQL Cancer Module [42], QOLCC [64, 65], SQOLPOP [66, 67]). Among the latter, one is specifically for children with advanced cancer (PAC-QoL), and another is for children with brain tumors (PedsQL Brain Tumor). Further characteristics are presented in Table 1.

Contents assessed by included PROMs

For all but one PROM (SQOLPOP), review copies or item lists were found. Four-hundred different items were retrieved, some of which belong to more than one length-version or module. Of these 400 items, 22 were excluded

as open-ended questions, determinant, or conditional items. No conflicts occurred in defining the question type.

The remaining 378 items were assigned to one of the domains, subdomains, and identifying concepts within the conceptual framework by Anthony et al. [6]. The reviewers agreed upon the categorization of 94.97% of items (359/378). The few conflicts were easily resolved, and the complementation of the HRQOL model for content categorization was discussed [MR, DR] (Additional file 3). The categorizations were adapted accordingly [MR], and the final categorization was approved again [DR].

Most items from the overall item pool cover psychological aspects. As displayed in Fig. 2, 35.19% (N=133) of items address emotional health and another 7.67% (N=19) refer to cognitive health. A quarter of items assess social (N=191, 26.72%) and physical health (N=89, 25.93%). Less than 5% measure general health perception or other aspects (i.e., financial).

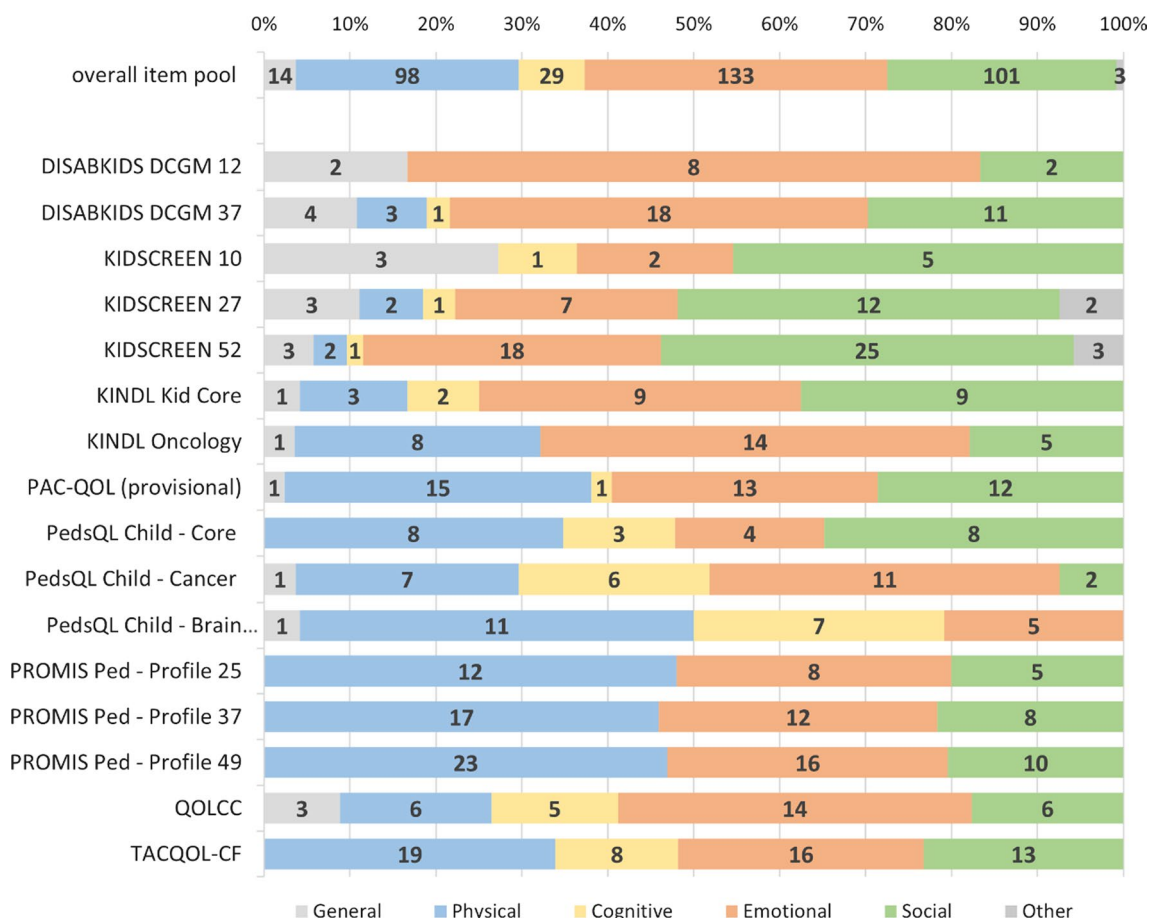


Fig. 2 Proportion and total number of items assessing the domains of health-related quality of life within the overall item pool and within the different questionnaires (Numbers in bars indicate the total number of items; length of bars indicates the proportion, compared to the legend above)

Upon closer inspection of the different PROMs (Fig. 2), it is apparent that the generic instruments and core scales (except for the PedsQL Generic Core Scale) assess less physical and more social issues than instruments designed for children with chronic diseases or cancer. In contrast, the PROMIS Pediatric Profile and the PedsQL Brain Tumor Module have the strongest focus on physical health, with approximately 50% of their items being dedicated to this domain. Cognitive issues are mostly represented in the PedsQL Brain Tumor and Cancer Modules, but not covered in the PROMIS Pediatric Profile. Additional file 4 provides more detail.

Quality ratings of development studies

The ratings obtained for the quality of development studies are displayed in Table 2, including justifications for ratings other than ‘very good’ (V). For most instruments, a clear definition of the construct to be measured, the target population, and the context was given. For the KINDL-R Oncology module, these points remained ‘doubtful’, as no development study was available. The SQOLPOP obtained an ‘inadequate’ rating, because the development study did not clarify which dimensions this questionnaire should capture [67].

The involvement of the target population in concept elicitation was rated ‘inadequate’ (five PROMs) or ‘doubtful’ (five PROMs) for most PROMs. In some cases, no children were involved in the development studies (PAC-QOL, SQOLPOP, TAC-QOL). For other PROMs, methods were described insufficiently. For example, for the PedsQL modules, it remains unclear how they were derived from the previous PCQL.

For four instruments, no cognitive interviews were conducted (KINDL-R Oncology, PedsQL Generic, PedsQL Cancer, TACQOL), in another three cases, it remained ‘doubtful’ whether they were conducted in the target population (PedsQL Brain Tumor, QOLCC-7-12, SQOLPOP). The remaining studies solely investigated comprehensibility, whereas comprehensiveness was often not investigated (DISABKIDS, KIDSCREEN, KINDL-R Generic, PAC-QOL). All but one had to be rated as ‘doubtful’ or even ‘inadequate’ for comprehensiveness, mostly because it remained unclear whether the identified difficulties were addressed and because items were not appropriately (re-)tested in their final form. The PROMIS Pediatric Profile was the only instrument, for which ‘very good’ methods were applied and reporting was good. Nevertheless, it received an ‘adequate’ rating only, because most items were tested in five or six patients, while a ‘very good’ rating would have required seven or more patients per item.

The total rating for the development was based on the quality of concept elicitation and the quality of cognitive

interview studies. The overall development was of ‘inadequate’ quality for eight PROMs and of ‘doubtful’ quality for another three PROMs. Only the PROMIS Pediatric Profile was informed by an ‘adequate’—almost ‘very good’—development procedure.

Quality ratings of content validity studies

Quality ratings for content validity studies are provided in Table 3, including justifications for ratings other than ‘very good’ (V). Content validity studies were only conducted for three PROMs, the DISABKIDS, the KINDL-R Generic Module, and the QOLCC-7-12. For all three, quality was rated ‘inadequate’. The QOLCC-7-12 was only evaluated with five healthcare-experts, but no patients or caregivers were involved [65, 100]. For the DISABKIDS, only a few written comments by children and parents were taken into account, while focus groups were held with nurses [55]. Furthermore, it is questionable whether the comments resulted in any adaptations. In the study investigating the KINDL-R Generic Module, children were asked to rate the relevance and comprehensibility of the whole questionnaire, but not for each item individually [76].

Rating of results and evidence grading

Following the COSMIN methodology, the development and content validity studies of mostly ‘doubtful’ or ‘inadequate’ quality can only provide ‘very low’ or ‘low’ evidence for the relevance, comprehensiveness, and comprehensibility of nearly all investigated PROMs. Only the PROMIS Pediatric Profile, with its ‘adequate’—almost ‘very good’—development procedure can rely on a ‘moderate’ evidence base for the three components of content validity. The quality of evidence for each PROM is displayed in Table 4, together with ratings of the results.

Due to the ‘very low’ evidence for most PROMs, the ratings often rely on reviewers’ ratings. As no review copy was available for the SQOLPOP, only ‘indeterminate’ ratings could be given for this instrument. For all other measures, ratings of results for relevance and comprehensiveness were based strictly on the content categorization described before. Relevance was rated as ‘sufficient’ because all items could be mapped onto the conceptual model of HRQOL. However, the comprehensiveness of seven PROMs was rated as ‘insufficient’, mostly because cognitive issues or positive psychological functioning were missing.

As all instruments have age-appropriate recall-periods and response-options, reviewers’ comprehensibility ratings were positive and/or followed the study results. Only for the KINDL-R Oncology Module, did reviewers rate the comprehensibility as ‘insufficient’, because its design is considerably complex. In

Table 2 Quality ratings of development studies following the COSMIN methodology

PROM	PROM development				Justification for Cognitive interviews (CI)				Total PROM development	
	PROM design and concept elicitation (CE)				Justification for given ratings other than 'very good' (V)					
General design requirements (GDR)		CE ¹	Total PROM design	Given ratings other than 'very good' (V)	Comprehensibility	Comprehensiveness	Total CI study	Justification for given ratings other than 'very good' (V)		
Construct	Origin of construct	Target population	Context of use	Developed in target population	GDR	Performed in target population				
DISABKIDS	V	V	V	V	D	D	D	D	D	Difficulties of children noticed; however, these were not considered in item reduction/reformulation process, which was only guided by quantitative data and expert consensus [52, 53, 69].
KIDSCREEN	V	V	V	V	V	V	I	I	I	Pilot-testing of 185 preliminary items resulted in adaptations; however, items were not tested in their final form and analysis and results were only described vaguely [48, 73].
KINDL-R Generic	V	V	V	V	D	V	D	D	D	Pre-test on comprehensibility of preliminary version conducted, but only healthy children (N = 28) and teachers, only written information, and revisions re-evaluated with only three healthy children [75].

Table 2 (continued)

PROM	PROM development					Total PROM design	Justification for Cognitive interviews (CI)				Total CI study	Justification for given ratings other than 'very good' (V)	Total PROM development		
	PROM design and concept elicitation (CE)						CE ¹	Justification for given ratings other than 'very good' (V)						Comprehensiveness	
	Construct	Origin of construct	Target population	Context of use	Developed in target population			Performed in target population	Comprehensibility	Comprehensiveness					
KINDL-R Oncology	D	D	V	D	I	N/A	I	N/A	N/A	N/A	I	No published development study available. When approached for literature on this module, the responsible office of QOL measures at the university clinic Hamburg-Eppendorf (UKE) referred to the study by Ergin et al. [60] on developing a Turkish translation (mail, 04/16/2020). However, this study does not report anything about the development of the original module.	I	No CI study available.	
PAC-QoL Child	V	V	V	V	I	N/A	I	D	N/A	D	D	I	No children were included in concept elicitation. For ethical reasons, Cataudella et al. [61] decided to only interview bereaved parents and HCPs at this stage.	I	CIs indicated that younger children had considerable difficulties in understanding, which has not been adequately addressed so far [62].

Table 2 (continued)

PROM	PROM development				CE ¹	Total PROM design	Justification for Cognitive interviews (CI)				Total CI study	Justification for given ratings other than 'very good' (V)	Total PROM development	
	PROM design and concept elicitation (CE)						Comprehensibility	Comprehensiveness	Performed in target population	Total CI study				
	General design requirements (GDR)	Origin of construct	Target population	Context of use										
PedsQL Generic	V	V	V	V	D	D	No clear information is available on how the PedsQL was derived from the previous PCOL [see also 31]. Brief descriptions of PCOL development do not provide sufficient information on sample and methods [79, 80, 82].	N/A	N/A	N/A	I	No CIs conducted for the latest version(s) of the PedsQL Generic Core Scale.		
PedsQL Cancer	V	V	V	V	D	D	Again, missing information on how the PedsQL was derived from PCOL [see also 31].	N/A	N/A	N/A	I	NO CI study available.	I	
PedsQL Brain Tumor	V	V	V	V	D	D	Palmer et al. [63] state that it was "developed through focus groups with healthcare providers, children and parents, cognitive interviews, pre-testing, and field testing protocols," (p. 288). Thus, it is assumable that children of the target population were involved. However, as no further information is given, the quality remains 'doubtful'.	D	D	D	D	D	Missing information results in 'doubtful' ratings, see statement by Palmer et al. [63] provided as justification for rating of development studies.	D

Table 2 (continued)

PROM	PROM development				Justification for Cognitive interviews (CI)				Total PROM development			
	PROM design and concept elicitation (CE)				Justification for given ratings other than 'very good' (V)	Total CI study	Justification for given ratings other than 'very good' (V)	Total PROM design				
	General design requirements (GDR)		Context of use							Comprehensibility	Siveness	
Construct	Origin of construct	Target population	Developed in target population	CE ¹	Justification for given ratings other than 'very good' (V)	Performed in target population	Comprehensibility	Siveness	Total CI study			Justification for given ratings other than 'very good' (V)
PROMIS Ped per scale Profile	V	V	V	V	V	V	A	A	A	Even though very good methods applied, rating off CIs can only be 'adequate', because 93% of the items have been tested in 5 or 6 patients only [85], which is insufficient to obtain a 'very good' rating in the COSMIN manual (7 required). Additional items developed by Quinn et al. [89] have been tested in at least 5 children per item.	A	
Anxiety and depression ²	"	"	"	"	"	"	"	"	"	CIs can only be 'adequate', because 93% of the items have been tested in 5 or 6 patients only [85], which is insufficient to obtain a 'very good' rating in the COSMIN manual (7 required). Additional items developed by Quinn et al. [89] have been tested in at least 5 children per item.	"	
Fatigue	"	"	"	"	"	"	"	"	"	Further item reduction relied on quantitative methods only (methods described in [103]; results per domain: [87, 90–93].	"	
Mobility and upper extremity ²	"	"	"	"	"	"	"	"	"	Only three children were included in CIs and no further information is available on methods or results [64].	"	
Pain intensity	"	"	"	"	"	"	"	"	"		"	
Pain interference	"	"	"	"	"	"	"	"	"		"	
Peer relation-ships	"	"	"	"	"	"	"	"	"		"	
QOLCC-7-12V	V	V	V	V	I	I	D	D	D	The qualitative methods applied were mostly 'adequate' to 'very good' [101]. 'Inadequate' rating because data were not rated by two independent investigators (Yeh was the sole author). However, it must be acknowledged that the results were discussed several times with faculty members, nursing students and participants, i.e., pediatric cancer patients.	D	I

Table 2 (continued)

PROM	PROM development					CE ¹	Justification for Cognitive interviews (CI)			Total PROM design	Total PROM development		
	PROM design and concept elicitation (CE)						Justification for given ratings other than 'very good' (V)	Comprehensibility	Comprehensiveness			Total CI study	
	Construct	Origin of construct	Target population	Context of use	Developed in target population								
SQOLPOP	I	V	V	V	I	N/A	I	D	D	D	D	I	Justification for given ratings other than 'very good' (V): Pilot-study conducted with 25 children and their parents, who apparently had no negative feedback to give [67]. However, missing information on sample and methods results in 'doubtful' rating.
TACQOL	V	V	V	V	I	N/A	I	N/A	N/A	N/A	I	I	Justification for given ratings other than 'very good' (V): No children involved in concept elicitation and issue generation. 'inadequate'. No children were involved in the concept elicitation.

V = very good; A = adequate; D = doubtful; I = inadequate; N/A = No study available; "see last rating above; bold = total ratings per section

¹ When the PROM was not developed in a sample representing the target population, the concept elicitation was not further rated

² PROMIS-scales developed together in one study are listed in combined rows

this PROM, some items require three responses: For symptoms, children must indicate frequency and the resulting burden. For treatment- or procedure-related issues, a conditional item is followed by frequency and burden ratings.

Discussion

The quality assessment of development, cognitive interview, and content validity studies showed that none of the investigated PROMs has a solid evidence base for its content validity. For most instruments, evidence is 'very low,' only the PROMIS Pediatric Profile is based on 'moderate' evidence. Overall, the scarce evidence available indicates that the PROMs cover relevant issues, while evidence for comprehensiveness and comprehensibility is partly inconsistent or indicates that these have not been sufficiently fulfilled.

Methodological shortcomings and possible explanations

The reasons for this low evidence level can be found in the study design, methodological quality, and insufficient reporting. As already stated by Klassen et al. [31], patients were not sufficiently involved. Guidelines on patient involvement in PROM development as well as reporting guidelines did only appear after most instruments had been developed. Thus, the developers of the investigated PROMs could not yet benefit from their guidance. The concept of content validity in particular has not been clearly defined for a long time.

Missing qualitative studies and patient involvement

Most of the PROMs were developed in the 1990s or early 2000s, before the publication of milestone policies by the European Medicines Agency (EMA) [108] and the American Food and Drug Administration (FDA) [109] and methodological guidelines on PROM development or content validity around 2010, e.g., by the International Society for Pharmacoeconomics and Outcomes Research Patient Reported Outcome Good Research Practices Task Force (ISPOR PRO) [24–26, 110] or the PROMIS developers [85, 86]. This might explain poor or inconsistent methods and reporting. However, missing or 'inadequate' development studies could be compensated by qualitative content validity studies to strengthen the evidence for existing tools. As an example, the content validity of the most widely used adult cancer questionnaire, the EORTC QLQ-C30, is currently being evaluated with adult [111] and adolescent cancer patients [112]. For the pediatric PROMs included in the present review, almost no content validity studies were available.

Lacking qualitative evidence, investigators take the mere use of questionnaires as an indicator of content validity. For example, Arabiat et al. state that "Face and content validity were assumed because the Ped-sQL™ (4.0) is widely used and reported in quality of life research" [83]. Despite strong recommendations for patient involvement, there are several barriers for qualitative research. Applying qualitative methods is partly a question of resources (i.e., financial means, infrastructure, collaborations, expertise, etc.). For example, Petersen et al., who interviewed children during the development procedure of the DISABKIDS, concluded that "these techniques are a helpful method. Nevertheless, the amount of time necessary to carry this out and analyze it is a weakness of this approach" [69]. Despite these challenges, qualitative methods are crucial, because content validity is a question of heuristics that cannot be resolved by quantitative methods.

Missing clarity about the concept of content validity

Another reason for missing research on content validity might be that this measurement property has been the subject of scientific dispute [113]. Following critique from modern test theory, guidelines seemingly struggled to redefine the concept and to identify methods for its assessment [113, 114]. It is only in the latest version of the COSMIN methodology that content validity is clearly described by the three components of relevance, comprehensiveness, and comprehensibility, and that corresponding standards and criteria are defined [21, 22]. This new and clear definition and the high requirements of the recent COSMIN guidelines make a considerable difference. Wayant et al. [35], who used the new methodology, found the same lack of evidence highlighted by our review. This is in contrast with reviews based on the older version, which came to very positive results [e.g., 34].

As the operationalization of content validity by relevance, comprehensibility, and comprehensiveness is still young, studies so far have seldom covered all three components separately and equally. For example, Kudubes and Bektas [67] asked health-care professionals only to rate how much change was needed for each item, without specifying what kind of change was required and why. If studies made a distinction between the three components, comprehensiveness was less often investigated compared to relevance and comprehensibility. This is in line with a recent review of studies on measurement properties of PROMs, which found that 77.8% of the studies assessed relevance, 48.2% evaluated comprehensibility, and only 3.7% focused on comprehensiveness [115].

When it comes to comprehensibility, there is again a lack of differentiation. Wayant et al. [35] state that

instructions were not investigated for any of the PROMs included in their review; rather, the studies focused solely on items. In our review, the PROMIS Pediatric Profile is the only tool for which items, instructions, response-options, and recall-periods were assessed separately [85]. For the KINDL Generic Module, which was developed a decade earlier, comprehensibility was not even rated per item, but for the whole questionnaire [76].

'Doubtful' ratings of study quality due to poor reporting

Not only is there a lack of qualitative studies of high quality for assessing content validity, but most 'doubtful' ratings were given due to insufficient reporting. In several cases, development and cognitive interview studies were only briefly described in a paragraph of a later study focusing on quantitative validity or reliability testing. Such shortcomings in reporting of qualitative methods in PROM development are a well-known problem and not specific to the field of pediatric oncology [116].

The recently published COSMIN reporting guideline will hopefully improve the situation [117]. However, it gives only very loose rules for content validity studies, defining *what* must be reported. It does not provide guidance on *how much detail* is required to meet the criteria of the COSMIN methodology for assessing content validity. Therefore, it might be useful to also have this methodology in mind when developing a new instrument. Even though Gagnier et al. differentiate clearly between the scopes of the two guidelines [117], it would surely help to prepare, conduct, and report future research more effectively and to provide more solid evidence.

Limitations and challenges of applying the COSMIN methodology on content validity assessment

We are aware that the search strategy underlying this review was limited. The search was conducted in only one database, PubMed, and did not rely on the extensive search filter by COSMIN [118]. This filter, however, is designed to find studies reporting all psychometric properties and not specifically content validity. Thus, the results would have exceeded the scope of our review. That no further PROMs could be identified through cross-checking with very comprehensive reviews [44, 45] indicates that our search was sufficiently fit for identifying relevant PROMs. Corresponding development and content validity studies are usually referred to as primary citations. Beyond that, we conducted additional searches and contacted PROM designers and authors to make sure that no relevant studies were missed.

While the COSMIN methodology is the current gold standard for assessing the quality criteria of PROMs, its application was partly challenging. Not only is the reporting inconsistent and insufficient, but the differentiation

between cognitive interview and content validity studies is sometimes difficult to make. Furthermore, the COSMIN guidelines propose rating each subscale separately [22]. This was rarely possible, because most of the multidimensional PROMs were developed as a whole and the information was not given per subscale. Even for the PROMIS Pediatric Profile, for which subscales were developed separately, not all steps and results were reported for each subscale in detail. These uncertainties led to many 'doubtful' ratings. Since the COSMIN methodology follows the worst-score-counts-principle, one 'doubtful' rating results in a 'doubtful' overall rating. This principle could be criticized for being too strict, as less relevant deficiencies could outweigh more important standards that were well met.

The situation is further complicated because the guidelines were not developed for pediatric tools and do not provide any advice on how to consider evidence provided by caregivers. We tried to resolve this by adding the standards required for expert involvement in content validity studies to take caregiver interviews into account. One could argue that caregivers' input should also have been considered in concept elicitation or cognitive interview studies. However, as caregiver- and patient-report often differ considerably, we decided to not systematically consider input from caregivers during these steps—in exactly the same way that the opinions of health-care professionals are ignored at this point following the COSMIN guidelines.

Conclusion and implications

Following the COSMIN methodology, this systematic review showed that there is only fragile evidence for the content validity of PROMs for HRQOL in children with cancer. Only the PROMIS Pediatric Profile has a 'moderate' level of evidence. Results indicate that it covers relevant issues and is comprehensible. Its comprehensiveness could be improved by adding further pediatric PROMIS scales (e.g., cognitive function, meaning and purpose, life satisfaction, positive affect) [43]. Thus, among the investigated PROMs, the Pediatric PROMIS Profile is recommended. However, this instrument is not disease-specific, and it might be worthwhile conducting a qualitative content validity study in children with cancer.

This lack of evidence can be explained by several factors: Most investigated instruments were developed before the publication of milestone policies and guidelines. Learning from the strengths and limitations of said previous PROM developments, these guidelines set new methodological standards. Content validity, in particular, was only clearly defined in the latest version of the COSMIN methodology. While it is, therefore,

Table 3 Quality ratings of content validity studies following the COSMIN methodology

PROM	Content validity studies						Justification for given ratings other than 'very good' (V)	
	Patients		Experts		Parents'			Total content validation
	Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness	Relevance		
DISABKIDS	I	I	I	D	D	I	I	Validity assessment of the DCGM-37 in cancer patients primarily based on quantitative methods; focus groups on 'feasibility' with nurses, only a few written comments by parents or patients; suggestions for reformulations and to add filter-questions did not result in according adaptations [55].
KIDSCREEN	N/A	N/A	N/A	N/A	N/A	N/A	I	The assessment of psychometric properties of the final 52-item-version were exclusively based on quantitative methods [47, 48]. For the development of the shorter versions with 27 and 10 items, no information was found.
KINDL-R Generic	I	I	I	I	I	N/A	I	A CV study in children with chronic diseases (asthma, diabetes, not cancer), was 'doubtful', as they rated the comprehensibility and relevance of the questionnaire as a whole instead of single items [76]. Further validation based on quantitative methods only [49, 50].
KINDL-R Oncology	N/A	N/A	N/A	N/A	N/A	N/A	I	As it relied solely on quantitative methods to analyze reliability and construct validity, the paper by Ergin et al. [60] can not be taken into account as a content validity study.
PAC-QoL Child	N/A	N/A	N/A	N/A	N/A	N/A	I	No CV study available, but within a short correspondence Dr. Cataudella (Research Gate, 08/17/2020) announced that another paper on face validity and preliminary psychometrics is in preparation.

Table 3 (continued)

PROM	Content validity studies						Justification for given ratings other than 'very good' (V)		
	Patients		Experts		Parents ¹			Total content validation	
	Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness	Relevance			Comprehensiveness
QOLCC-7-12	I	I	I	D	D	N/A	N/A	I	CV was assessed within a Master Thesis, which is only available in Chinese [Huang, 2000 cited in 100]. Following scarce information provided in later studies [65, 100], no patients or parents were included. Five experts were involved to rate the relevance, analyzed with the content validity index (CVI). As it is unclear whether two researchers conducted the analysis, the quality remains 'doubtful'. Further validity and reliability testing by Kudubes and Bektas [67] relies on quantitative methods only and therefore cannot be considered as CV study. No other CV study available.
SOQLPOP	N/A	N/A	N/A	N/A	N/A	N/A	N/A	I	No CV study available. The reference study relied on quantitative methods only [102].
TACQOL	N/A	N/A	N/A	N/A	N/A	N/A	N/A	I	No CV study available. The reference study relied on quantitative methods only [102].

V, very good; A, adequate; D, doubtful; I, inadequate; N/A, no study available; " see last rating above; bold = total ratings per section

¹ This section was added to take the central role of parents in pediatric healthcare into account. As it is not required by the COSMIN Guidelines, 'inadequate' ratings were only given when parents were involved using 'inadequate' methods. If content validity studies were performed without parents, parent involvement was not rated

² PROMIS-scales developed together in one study are listed in combined rows

Table 4 Evidence grading and overall ratings for the relevance, comprehensiveness, and comprehensibility of the included patient-reported outcome measures (PROMs) for health-related quality of life (HRQOL) assessment in children with cancer

PROM	Relevance			Comprehensiveness			Comprehensibility		
	Rating of Results	Justification for rating of results	Quality of evidence	Rating of results	Justification for rating of results	Quality of evidence	Rating of results	Justification for rating of results	Quality of evidence
DISABKIDS 12/37	+	No cancer patients involved, but all items reflect concepts of the conceptual model for QOL provided by Anthony et al. [6].	Low	–	Mapping indicates under-representation of physical and cognitive issues (DCGM-12: 0 physical, 0 cognitive items; DCGM-37: 3 physical, 1 cognitive items).	Low	–	Items, response options, and recall period are age-appropriate; but unclear if/how difficulties identified in studies were addressed [55, 69].	Low
KIDSCREEN 10/27/52	+	More than 90% of the items (except items on financial situation: KIDSCREEN 27: 2 items, KIDSCREEN 52: 3 items) match concepts of the QOL model by Anthony et al. [6].	Very low	–	Mapping indicates under-representation of physical and cognitive issues. KIDSCREEN 10, 27, 52 cover 0, 2, 2 physical and 1, 1, 1 cognitive items, respectively. Core symptoms like pain or nausea are missing	Very low	+	Studies state that no difficulties [48, 73]. This impression is shared by the reviewers.	Very low
Kid KINDL-R Generic	+	When asked about the relevance of the whole questionnaire, 80% report perceiving it as very relevant or relevant [76]. All items could be mapped onto the model of QOL by Anthony et al. [6].	Low	–	Mapping indicates under-representation of physical (3 items) and cognitive (2 items) issues in the generic tool, which might explain why 20% of chronically ill children did not perceive the questionnaire as a whole as relevant [76].	Low	+	“95.5% of the chronically ill children found the questions easily understandable” [76].	Low

Table 4 (continued)

PROM	Relevance		Comprehensiveness		Comprehensibility				
	Rating of Results	Justification for rating of results	Quality of evidence	Rating of results	Justification for rating of results	Quality of evidence	Rating of results	Justification for rating of results	Quality of evidence
Kid-Kiddo KINDL-R Oncology	+	All items could be mapped onto the QOL model by Anthony et al. [6].	Very low	-	Mapping indicates that the tool is not comprehensive. Cognitive issues are not assessed at all. From the psychological domain, only treatment burden ("bothered"; 12 items) and moodiness (2 items) are assessed. Neither other forms emotional distress nor positive mental health aspects are covered.	Very low	-	The reviewers rated the comprehensibility as insufficient due to very complex response-options. Many items require more than one response: For symptoms, children must indicate frequency and the resulting burden. For treatment- or procedure-related issues, a conditional item is followed by a frequency and a burden rating.	Very low
PAC-QoL Child	+	Items mapped well onto the conceptual model by Anthony et al. [6], except four items covering involvement in care, which was added as a new subdomain in this review.	Very low	+	All domains and nearly all subdomains (except self-esteem and treatment-burden) are covered.	Very low	-	Results of studies available at the time indicate insufficient comprehensibility [62].	Very low
PedsQL Generic Core Scale 4.0	+	All items cover issues reflecting the QOL model by Anthony et al. [6].	Very low	-	No items assess positive psychological functioning, self-esteem, or body image.	Very low	+	Reviewers' ratings indicate sufficient comprehensibility because instructions, items, and response-options are considered age-appropriate.	Very low

Table 4 (continued)

PROM	Relevance		Comprehensiveness		Comprehensibility			
	Rating of Results	Justification for rating of results	Quality of evidence	Rating of results	Justification for rating of results	Quality of evidence		
PedsQL Cancer Module 3.0	+	All items cover issues reflecting the QOL model by Anthony et al. [6].	Very low	-	No items on positive psychological functioning are assessed. The social domain does only contain items on involvement in care, whereas the social context of family, peers, school, and leisure time is missing.	Very low	Reviewers' ratings indicate sufficient comprehensibility because instructions, items, and response options are considered age-appropriate.	Very low
PedsQL Brain Tumor Module	+	All items cover issues reflecting the QOL model by Anthony et al. [6].	Very low ¹	-	No items on positive psychological functioning or social domain.	Very low ¹	Reviewers' ratings indicate sufficient comprehensibility because instructions, items, and response options are considered age-appropriate.	Very low ¹
PROMIS Ped Profile 25/37/49	+	All items cover issues reflecting the QOL model by Anthony et al. [6]. Jones et al. [105] did not report any general relevance issues when comparing child- and parent ratings on relevance.	Moderate	±	Healthy children and pediatric asthma patients had nothing to add [85]. However, this might not be representative for children with cancer. (Children with chronic kidney or Crohn's disease and adult cancer patients had disease-specific issues to add; see [106, 107]). Mapping indicates that cognitive as well as positive psychological issues are missing. Even though the physical domain takes approximately the half of items, none covers nausea, a core symptom in oncology.	Moderate	PROMIS is the only inventory reporting comprehensibility for items, response options, and instructions separately and results indicate very good comprehensibility [85].	Moderate

Table 4 (continued)

PROM	Relevance		Comprehensiveness		Comprehensibility	
	Rating of Results	Justification for rating of results	Quality of evidence	Rating of results	Justification for rating of results	Quality of evidence
QOLCC-7-12	+	Yeh et al. [64] report that besides quantitative criteria, "items were retained if at least two pediatric oncology specialists or three patients from the qualitative study considered them clinically important" (p.163). Most items mapped well onto the model by Anthony et al. [6], except three items on knowledge about cancer and three items on involvement in care—aspects which were added as new subdomains.	Very low	+	Mapping indicates that all relevant domains are represented, even though the physical domain takes a smaller proportion than in other cancer-specific tools.	Very low
SQLPOP	?	In case of 'very low' evidence, the rating of results should rely on reviewers' ratings only. However, as no review copy or item list was available, it was impossible to make any ratings.	Very low	?	In case of 'very low' evidence, the rating of results should rely on reviewers' ratings only. However, as no review copy or item list was available, it was impossible to make any ratings.	Very low
TACQOL	+	All items cover issues reflecting the QOL model by Anthony et al. [6].	Very low	+	All domains of the QOL model by Anthony et al. [6] are covered and take a reasonable proportion.	Very low

+ , sufficient; - , insufficient; ± , inconsistent; ? , indeterminate (only if no evidence from studies and no access to review copy for own ratings)
 VL = very low (PROM development study of 'inadequate' quality AND only content validity studies of 'inadequate' quality or no content validity studies (only reviewers' ratings))
 L = low (PROM development study of 'doubtful' quality AND only content validity studies of 'inadequate' quality or no content validity studies)
 M = moderate (PROM development study of 'very good' or 'adequate' quality AND only content validity studies of 'inadequate' quality or no content validity studies; OR at least one content validity study of 'doubtful' quality)
 1 Following the COSMIN guidelines, evidence was 'low', because the development procedure is 'doubtful' and no content validity study is available. However, no study results are available, so that rating of results is solely based on reviewers' ratings. Thus, the evidence was down-graded to 'very low'. This is in line with a current review by Bull et al. [37], who state that no information on the content validity is provided for the PedsQL Brain Tumor Module.

understandable that previous projects did not fulfill all required standards, PRO and HRQOL research in pediatric oncology should still try to catch up with the scientific and methodological progress of the last decade.

Therefore, we argue that further efforts are needed to provide PROMs for HRQOL assessment in children with cancer that are based on solid evidence. This could include the development of new instruments, as well as performing content validity studies to strengthen the evidence for already-existing PROMs. In each case, it is strongly recommended that existing guidelines on qualitative methods and reporting standards for these study types be adhered to. Within the EORTC QLQ, we are currently developing an HRQOL questionnaire for children with cancer [119]. Following the EORTC QLQ module development guidelines [23], this involves not only a literature review [45], but also in-depth interviews with children with cancer, their parents, and health-care professionals.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41687-023-00540-8>.

Additional file 1. PRISMA Checklist.

Additional file 2. Categorization Rules.

Additional file 3. Adaptations to Complement the Model of Health-Related Quality of Life.

Additional file 4. List of items per Patient-Reported Outcome Measure (PROM) covering the various domains, subdomains and identifying concepts of health-related quality of life.

Acknowledgements

We would like to thank Prof. Caroline Terwee, co-founder of the COSMIN initiative, for her advice on adapting the COSMIN methodology to include caregivers' input. Beyond that, we thank Dr. Luz Fialho, director of outcome research at the International Consortium for Health Outcomes Measurement (ICHOM), for kindly sharing the list of Patient-Reported Outcome Measures (PROMs) collected for the development of the Overall Pediatric Health Standard Set (OPH-SS).

Author contributions

The present paper is based on Maria Rothmund's master thesis, supervised by Gerhard Rumpold and David Riedl and submitted to the Institute of Psychology at the University of Innsbruck, Austria. Conceptualization: MR, GR, DR; Methodology: MR, AM, DR; Formal analysis and investigation: MR, AM, DR; Writing—original draft preparation: MR; Writing—review and editing: MR, AM, GR, RC, SS, ASD, DR; Funding acquisition: SS, DR; Supervision: GR, SS, ASD, DR. All authors read and approved the final manuscript.

Funding

This study was supported by the European Organisation for Research and Treatment of Cancer Quality of Life Group (EORTC QLQ). The grant (no. 002–2020) was awarded to David Riedl and Samantha Sodergren.

Availability of data and materials

The authors declare that the relevant data supporting our findings is provided in the article and the supplementary files. For further requests, please contact the corresponding author.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors collaborate in projects developing new HRQOL questionnaires for children and adolescents and young adults (AYAs) with cancer on behalf of the EORTC QLQ.

Author details

¹Department of Psychiatry, Psychotherapy, Psychosomatics and Medical Psychology, University Clinic of Psychiatry II, Medical University Innsbruck, Innsbruck, Austria. ²Institute of Psychology, University of Innsbruck, Innsbruck, Austria. ³Department of Pediatrics I, Medical University Innsbruck, Innsbruck, Austria. ⁴School of Health Sciences, University of Southampton, Southampton, UK. ⁵Ludwig Boltzmann Institute for Rehabilitation Research, Vienna, Austria.

Received: 8 August 2022 Accepted: 3 January 2023

Published online: 19 January 2023

References

- American Cancer Society (2020) Cancer facts & figures 2020. American Cancer Society, Atlanta
- Erdmann F, Frederiksen LE, Bonaventure A et al (2020) Childhood cancer: survival, treatment modalities, late effects and improvements over time. *Cancer Epidemiol*. <https://doi.org/10.1016/j.canep.2020.101733>
- Gatta G, Botta L, Rossi S et al (2014) Childhood cancer survival in Europe 1999–2007: results of EUROCare-5—a population-based study. *Lancet Oncol* 15:35–47. [https://doi.org/10.1016/S1470-2045\(13\)70548-5](https://doi.org/10.1016/S1470-2045(13)70548-5)
- Driscoll JJ, Rixe O (2009) Overall survival: still the gold standard: why overall survival remains the definitive end point in cancer clinical trials. *Cancer J* 15:401–405. <https://doi.org/10.1097/PPO.0b013e3181bdc2e0>
- WHOQOL Group (1995) The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. *Soc Sci Med* 41:1403–1409
- Anthony SJ, Selkirk E, Sung L et al (2014) Considering quality of life for children with cancer: a systematic review of patient-reported outcome measures and the development of a conceptual model. *Qual Life Res* 23:771–789. <https://doi.org/10.1007/s11136-013-0482-x>
- Mack JW, McFatrigh M, Withycombe JS et al (2020) Agreement between child self-report and caregiver-proxy report for symptoms and functioning of children undergoing cancer treatment. *JAMA Pediatr*. <https://doi.org/10.1001/jamapediatrics.2020.2861>
- Baggott C, Cooper BA, Marina N et al (2014) Symptom assessment in pediatric oncology: how should concordance between children's and parents' reports be evaluated? *Cancer Nurs* 37:252–262. <https://doi.org/10.1097/NCC.0000000000000111>
- Parsons SK, Fairclough DL, Wang J et al (2012) Comparing longitudinal assessments of quality of life by patient and parent in newly diagnosed children with cancer: the value of both raters' perspectives. *Qual Life Res* 21:915–923. <https://doi.org/10.1007/s11136-011-9986-4>
- Varni JW, Thissen D, Stucky BD et al (2015) Item-level informant discrepancies between children and their parents on the PROMIS[®] pediatric scales. *Qual Life Res* 24:1921–1937. <https://doi.org/10.1007/s11136-014-0914-2>
- Chang P-C, Yeh C-H (2005) Agreement between child self-report and parent proxy-report to evaluate quality of life in children with cancer. *Psychooncology* 14:125–134. <https://doi.org/10.1002/pon.828>
- Yoo H-J, Ra Y-S, Park H-J et al (2010) Agreement between pediatric brain tumor patients and parent proxy reports regarding the Pediatric Functional Assessment of Cancer Therapy-Childhood Brain Tumor Survivors questionnaire, version 2. *Cancer* 116:3674–3682. <https://doi.org/10.1002/ncr.25200>

13. Riley AW (2004) Evidence that school-age children can self-report on their health. *Ambul Pediatr* 4:371–376. <https://doi.org/10.1367/A03-178R1>
14. Varni JW, Limbers CA, Burwinkle TM (2007) 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 4.0 Generic Core Scales. *Health Qual Life Outcomes* 5:1. <https://doi.org/10.1186/1477-7525-5-1>
15. Arbuckle R, Abetz-Webb L (2013) “Not just little adults”: qualitative methods to support the development of pediatric patient-reported outcomes. *Patient* 6:143–159. <https://doi.org/10.1007/s40271-013-0022-3>
16. Leahy AB, Steineck A (2020) Patient-reported outcomes in pediatric oncology: the patient voice as a gold standard. *JAMA Pediatr*. <https://doi.org/10.1001/jamapediatrics.2020.2868>
17. Graham A, Powell M, Taylor N et al (2013) Ethical research involving children. <https://childethics.com/wp-content/uploads/2013/10/ERIC-compendium-approved-digital-web.pdf>. Accessed 10 January 2023
18. Coyne I, Amory A, Gibson F et al (2016) Information-sharing between healthcare professionals, parents and children with cancer: more than a matter of information exchange. *Eur J Cancer Care (England)* 25:141–156. <https://doi.org/10.1111/ecc.12411>
19. Zwaanswijk M, Tates K, van Dulmen S et al (2007) Young patients’, parents’, and survivors’ communication preferences in paediatric oncology: results of online focus groups. *BMC Pediatr* 7:35. <https://doi.org/10.1186/1471-2431-7-35>
20. Mokkink LB, Terwee CB, Patrick DL et al (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 63:737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
21. Prinsen CAC, Mokkink LB, Bouter LM et al (2018) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 27:1147–1157. <https://doi.org/10.1007/s11136-018-1798-3>
22. Terwee CB, Prinsen CAC, Chiarotto A et al (2018) COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 27:1159–1170. <https://doi.org/10.1007/s11136-018-1829-0>
23. Wheelwright S, Bjordal K, Bottomley A et al (2021) EORTC quality of life group guidelines for developing questionnaire modules, 5th edn. <https://www.eortc.org/app/uploads/sites/2/2022/07/Module-Guide-lines-Version-5-FINAL.pdf>. Accessed 10 January 2023
24. Matza LS, Patrick DL, Riley AW et al (2013) Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health* 16:461–479. <https://doi.org/10.1016/j.jval.2013.04.004>
25. Patrick DL, Burke LB, Gwaltney CJ et al (2011) Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. *Value Health* 14:967–977. <https://doi.org/10.1016/j.jval.2011.06.014>
26. Patrick DL, Burke LB, Gwaltney CJ et al (2011) Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. *Value Health* 14:978–988. <https://doi.org/10.1016/j.jval.2011.06.013>
27. Coombes L, Bristowe K, Ellis-Smith C et al (2021) Enhancing validity, reliability and participation in self-reported health outcome measurement for children and young people: a systematic review of recall period, response scale format, and administration modality. *Qual Life Res*. <https://doi.org/10.1007/s11136-021-02814-4>
28. Withycombe JS, McFatrach M, Pinheiro L et al (2019) The association of age, literacy, and race on completing patient-reported outcome measures in pediatric oncology. *Qual Life Res* 28:1793–1801. <https://doi.org/10.1007/s11136-019-02109-9>
29. Sodergren SC, Husson O, Robinson J et al (2017) Systematic review of the health-related quality of life issues facing adolescents and young adults with cancer. *Qual Life Res* 26:1659–1672. <https://doi.org/10.1007/s11136-017-1520-x>
30. Sodergren SC, Husson O, Rohde GE et al (2018) A life put on pause: an exploration of the health-related quality of life issues relevant to adolescents and young adults with cancer. *J Adolesc Young Adult Oncol* 7:453–464. <https://doi.org/10.1089/jayao.2017.0110>
31. Klassen AF, Strohm SJ, Maurice-Stam H et al (2010) Quality of life questionnaires for children with cancer and childhood cancer survivors: a review of the development of available measures. *Support Care Cancer* 18:1207–1217. <https://doi.org/10.1007/s00520-009-0751-y>
32. Anthony SJ, Selkirk E, Sung L et al (2017) Quality of life of pediatric oncology patients: do patient-reported outcome instruments measure what matters to patients? *Qual Life Res* 26:273–281. <https://doi.org/10.1007/s11136-016-1393-4>
33. Hinds PS, Gattuso JS, Fletcher A et al (2004) Quality of life as conveyed by pediatric patients with cancer. *Qual Life Res* 13:761–772
34. Coombes LH, Wiseman T, Lucas G et al (2016) Health-related quality-of-life outcome measures in paediatric palliative care: a systematic review of psychometric properties and feasibility of use. *Palliat Med* 30:935–949. <https://doi.org/10.1177/0269216316649155>
35. Wayant C, Bixler K, Garrett M et al (2022) Evaluation of patient-reported outcome measures of positive psychosocial constructs in children and adolescent/young adults with cancer: a systematic review of measurement properties. *J Adolesc Young Adult Oncol* 11:78–94. <https://doi.org/10.1089/jayao.2021.0031>
36. Pinheiro LC, McFatrach M, Lucas N et al (2018) Child and adolescent self-report symptom measurement in pediatric oncology research: a systematic literature review. *Qual Life Res* 27:291–319. <https://doi.org/10.1007/s11136-017-1692-4>
37. Bull KS, Hornsey S, Kennedy CR et al (2020) Systematic review: measurement properties of patient-reported outcome measures evaluated with childhood brain tumor survivors or other acquired brain injury. *Neurooncol Pract* 7:277–287. <https://doi.org/10.1093/nop/npz064>
38. Mokkink LB, Terwee CB, Patrick DL et al (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 19:539–549. <https://doi.org/10.1007/s11136-010-9606-8>
39. Mokkink LB, Terwee CB, Knol DL et al (2010) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol*. <https://doi.org/10.1186/1471-2288-10-22>
40. Terwee CB, Mokkink LB, Knol DL et al (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 21:651–657. <https://doi.org/10.1007/s11136-011-9960-1>
41. Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. <https://doi.org/10.1136/bmj.n71>
42. Varni JW, Burwinkle TM, Katz ER et al (2002) The PedsQL in pediatric cancer: reliability and validity of the Pediatric Quality of Life Inventory generic core scale, multidimensional fatigue scale, and cancer module. *Cancer* 94:2090–2106
43. HealthMeasures (2021) List of pediatric measures. <https://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/list-of-pediatric-measures>. Accessed 30 June 2021
44. Algurén B, Ramirez JP, Salt M et al (2020) Development of an international standard set of patient-centred outcome measures for overall paediatric health: a consensus process. *Arch Dis Child*. <https://doi.org/10.1136/archdischild-2020-320345>
45. Rothmund M, Sodergren SC, Rohde GE et al (2022) Updating our understanding of health-related quality of life issues in children with cancer: a systematic review of patient-reported outcome measures and qualitative studies. *Qual Life Res*. <https://doi.org/10.1007/s11136-022-03259-z>
46. Mokkink LB, Prinsen CAC, Patrick DL et al (2018) COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs): User Manual. Version 1.0. <https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf>. Accessed 24 May 2021

47. Kidscreen Group Europe (2016) The Kidscreen questionnaires: Quality of life questionnaires for children and adolescents—handbook, 3rd edn. Pabst Science Publishers, Lengerich
48. Ravens-Sieberer U, Gosch A, Rajmil L et al (2005) KIDSCREEN-52 quality-of-life measure for children and adolescents. *Expert Rev Pharmacoecon Outcomes Res* 5:353–364. <https://doi.org/10.1586/14737167.5.3.353>
49. Ravens-Sieberer U, Bullinger M (2000) KINDL-R: questionnaire for measuring health-related quality of life in children and adolescents—revised version. Manual. <https://www.kindl.org/deutsch/sprachversionen/englisch/>. Accessed 13 May 2021
50. Bullinger M, Brütt AL, Erhart M et al (2008) Psychometric properties of the KINDL-R questionnaire: results of the BELLA study. *Eur Child Adolesc Psychiatry* 17(Suppl 1):125–132. <https://doi.org/10.1007/s00787-008-1014-z>
51. Varni JW, Seid M, Kurtin PS (2001) PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Med Care* 39:800–812
52. Muehlan H (2010) Developing the DCGM-12: a short-form of the DISAB-KIDS condition-generic module assessing health related quality of life in children and adolescents with chronic conditions. Doctoral Thesis, University of Hamburg
53. Simeoni M-C, Schmidt S, Muehlan H et al (2007) 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* 16:881–893. <https://doi.org/10.1007/s11136-007-9188-2>
54. The European DISABKIDS Group (2004) The DISABKIDS questionnaires. Quality of life questionnaires for children with chronic conditions: manual
55. af Sandeberg M, Johansson EM, Hagell P et al (2010) Psychometric properties of the DISABKIDS Chronic Generic Module (DCGM-37) when used in children undergoing treatment for cancer. *Health Qual Life Outcomes* 8:109. <https://doi.org/10.1186/1477-7525-8-109>
56. Hinds PS, Nuss SL, Ruccione KS et al (2013) PROMIS pediatric measures in pediatric oncology: valid and clinically feasible indicators of patient-reported outcomes. *Pediatr Blood Cancer* 60:402–408. <https://doi.org/10.1002/pbc.24233>
57. Hinds PS, Wang J, Cheng YI et al (2019) PROMIS pediatric measures validated in a longitudinal study design in pediatric oncology. *Pediatr Blood Cancer* 66:e27606. <https://doi.org/10.1002/pbc.27606>
58. Vogels T, Verrips GHW, Verloove-Vanhorick SP et al (1998) Measuring health-related quality of life in children: the development of the TAC-QOL parent form. *Qual Life Res* 7:457–465. <https://doi.org/10.1023/A:1008848218806>
59. Vogels T, Verrips GHW, Koopman HM et al (2000) TACQOL Manual. Parent and Child Form. <https://publications.tno.nl/publication/34636899/ffBvk3/vogels-2000-tacqolmanual.pdf>. Accessed 10 January 2023
60. Ergin D, Eser E, Kantar M et al (2015) Psychometric properties of the oncology module of the KINDL scale: first results. *J Pediatr Oncol Nurs* 32:83–95. <https://doi.org/10.1177/1043454214543020>
61. Cataudella D, Morley TE, Nesin A et al (2014) Development of a quality of life instrument for children with advanced cancer: the pediatric advanced care quality of life scale (PAC-QoL). *Pediatr Blood Cancer* 61:1840–1845. <https://doi.org/10.1002/pbc.25115>
62. Morley TE, Cataudella D, Fernandez CV et al (2014) Development of the Pediatric Advanced Care Quality of Life Scale (PAC-QoL): evaluating comprehension of items and response options. *Pediatr Blood Cancer* 61:1835–1839. <https://doi.org/10.1002/pbc.25111>
63. Palmer SN, Meeske KA, Katz ER et al (2007) The PedsQL Brain Tumor Module: initial reliability and validity. *Pediatr Blood Cancer* 49:287–293. <https://doi.org/10.1002/pbc.21026>
64. Yeh C-H, Chao K-Y, Hung L-C (2004) The quality of life for cancer children (QOLCC) in Taiwan (part I): reliability and construct validity by confirmatory factor analysis. *Psychooncology* 13:161–170. <https://doi.org/10.1002/pon.728>
65. Yeh C-H, Hung L-C, Chao K-Y (2004) The quality of life for cancer children (QOLCC) for Taiwanese children with cancer (part II): feasibility, cross-informants variance and clinical validity. *Psychooncology* 13:171–176. <https://doi.org/10.1002/pon.729>
66. Bektas M, Akdeniz Kudubes A, Ugur O et al (2016) Developing the scale for quality of life in pediatric oncology patients aged 13–18: adolescent form and parent form. *Asian Nurs Res (Korean Soc Nurs Sci)* 10:106–115. <https://doi.org/10.1016/j.anr.2016.03.002>
67. Kudubes AA, Bektas M (2015) Developing a scale for quality of life in pediatric oncology patients aged 7–12—children and parent forms. *Asian Pac J Cancer Prev* 16:523–529. <https://doi.org/10.7314/apjcp.2015.16.2.523>
68. Bullinger M, Schmidt S, Petersen C et al (2002) Assessing quality of life of children with chronic health conditions and disabilities: a European approach. *Int J Rehabil Res* 25:197–296
69. Petersen C, Schmidt S, Power M et al (2005) 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* 14:1065–1077. <https://doi.org/10.1007/s11136-004-2575-z>
70. Herdman M, Ravens-Sieberer U, Bullinger M et al (2002) Expert consensus in the development of a European health-related quality of life measure for children and adolescents: a Delphi study. *Acta Paediatr* 91:1385–1390
71. Baars RM, Atherton CI, Koopman HM et al (2005) The European DISABKIDS project: development of seven condition-specific modules to measure health related quality of life in children and adolescents. *Health Qual Life Outcomes* 3:70. <https://doi.org/10.1186/1477-7525-3-70>
72. Ravens-Sieberer U, Schmidt S, Gosch A et al (2007) Measuring subjective health in children and adolescents: results of the European KIDSCREEN/DISABKIDS Project. *German Med Sci* 4:Doc08
73. Detmar SB, Bruil J, Ravens-Sieberer U et al (2006) The use of focus groups in the development of the KIDSCREEN HRQL questionnaire. *Qual Life Res* 15:1345–1353. <https://doi.org/10.1007/s11136-006-0022-z>
74. Ravens-Sieberer U, Gosch A, Abel T et al (2001) Quality of life in children and adolescents: a European public health perspective. *Soz Präventivmed* 46:294–302
75. Bullinger M, von Mackensen S, Kirchberger I (1994) KINDL - ein Fragebogen zur Erfassung der gesundheitsbezogenen Lebensqualität von Kindern. *Zeitschrift für Gesundheitspsychologie* 2:64–77
76. Ravens-Sieberer U, Bullinger M (1998) Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results. *Qual Life Res* 7:399–407. <https://doi.org/10.1023/a:1008853819715>
77. Bullinger M, Ravens-Sieberer U (2006) Lebensqualität und chronische Krankheit: die Perspektive von Kindern und Jugendlichen in der Rehabilitation. *Prax Kinderpsychol Kinderpsychiatr* 55:23–35
78. Seid M, Varni JW, Rode CA et al (1999) The Pediatric Cancer Quality of Life Inventory: a modular approach to measuring health-related quality of life in children with cancer. *Int J Cancer* 83:71–76
79. Varni JW, Katz ER, Seid M et al (1998) The pediatric cancer quality of life inventory-32 (PCQL-32): I. Reliability Validity *Cancer* 82:1184–1196
80. Varni JW, Katz ER, Seid M et al (1998) The Pediatric Cancer Quality of Life Inventory (PCQL). I. Instrument development, descriptive statistics, and cross-informant variance. *J Behav Med* 21:179–204
81. Caru M, Perreault S, Levesque A et al (2021) Validity and reliability of the French version of the Pediatric Quality of Life Inventory™ brain tumor module. *Qual Life Res*. <https://doi.org/10.1007/s11136-021-02815-3>
82. Varni JW, Seid M, Rode CA (1999) The PedsQL™: measurement model for the Pediatric Quality of Life Inventory. *Med Care* 37:126–139
83. Arabiat D, Elliott B, Draper P et al (2011) Cross-cultural validation of the Pediatric Quality of Life Inventory™ 4.0 (PedsQL™) generic core scale into Arabic language. *Scand J Caring Sci* 25:828–833. <https://doi.org/10.1111/j.1471-6712.2011.00889.x>
84. Lau JTF, Yu X-n, Chu Y et al (2010) Validation of the Chinese version of the Pediatric Quality of Life Inventory (PedsQL) cancer module. *J Pediatr Psychol* 35:99–109. <https://doi.org/10.1093/jpepsy/jsp035>
85. Irwin DE, Varni JW, Yeatts K et al (2009) Cognitive interviewing methodology in the development of a pediatric item bank: a patient reported outcomes measurement information system (PROMIS) study. *Health Qual Life Outcomes* 7:3. <https://doi.org/10.1186/1477-7525-7-3>
86. Walsh TR, Irwin DE, Meier A et al (2008) The use of focus groups in the development of the PROMIS pediatrics item bank. *Qual Life Res* 17:725–735. <https://doi.org/10.1007/s11136-008-9338-1>

87. Irwin DE, Stucky B, Langer MM et al (2010) An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Qual Life Res* 19:595–607. <https://doi.org/10.1007/s11136-010-9619-3>
88. Irwin DE, Stucky BD, Thissen D et al (2010) Sampling plan and patient characteristics of the PROMIS pediatrics large-scale survey. *Qual Life Res* 19:585–594. <https://doi.org/10.1007/s11136-010-9618-4>
89. Quinn H, Thissen D, Liu Y et al (2014) Using item response theory to enrich and expand the PROMIS pediatric self report banks. *Health Qual Life Outcomes* 12:1–10
90. DeWalt DA, Thissen D, Stucky BD et al (2013) PROMIS Pediatric Peer Relationships Scale: development of a peer relationships item bank as part of social health measurement. *Health Psychol* 32:1093–1103. <https://doi.org/10.1037/a0032670>
91. DeWitt EM, Stucky BD, Thissen D et al (2011) Construction of the eight-item Patient-Reported Outcomes Measurement Information System pediatric physical function scales: built using item response theory. *J Clin Epidemiol* 64:794–804. <https://doi.org/10.1016/j.jclinepi.2010.10.012>
92. Lai J-S, Stucky BD, Thissen D et al (2013) Development and psychometric properties of the PROMIS[®] pediatric fatigue item banks. *Qual Life Res* 22:2417–2427. <https://doi.org/10.1007/s11136-013-0357-1>
93. Varni JW, Stucky BD, Thissen D et al (2010) PROMIS Pediatric Pain Interference Scale: an item response theory analysis of the pediatric pain item bank. *J Pain* 11:1109–1119. <https://doi.org/10.1016/j.jpain.2010.02.005>
94. Lai J-S, Kupst MJ, Beaumont JL et al (2019) Using the Patient-Reported Outcomes Measurement Information System (PROMIS) to measure symptom burden reported by patients with brain tumors. *Pediatr Blood Cancer* 66:e27526. <https://doi.org/10.1002/psc.27526>
95. Liu Y, Yuan C, Wang J et al (2016) Comparability of the Patient-Reported Outcomes Measurement Information System Pediatric short form symptom measures across culture: examination between Chinese and American children with cancer. *Qual Life Res* 25:2523–2533. <https://doi.org/10.1007/s11136-016-1312-8>
96. Menard JC, Hinds PS, Jacobs SS et al (2014) Feasibility and acceptability of the Patient-Reported Outcomes Measurement Information System measures in children and adolescents in active cancer treatment and survivorship. *Cancer Nurs* 37:66–74. <https://doi.org/10.1097/NCC.0b013e3182a0e23d>
97. Reeve BB, Edwards LJ, Jaeger BC et al (2018) Assessing responsiveness over time of the PROMIS[®] pediatric symptom and function measures in cancer, nephrotic syndrome, and sickle cell disease. *Qual Life Res* 27:249–257. <https://doi.org/10.1007/s11136-017-1697-z>
98. Westmoreland K, Reeve BB, Amugandoh A et al (2018) Translation, psychometric validation, and baseline results of the Patient-Reported Outcomes Measurement Information System (PROMIS) pediatric measures to assess health-related quality of life of patients with pediatric lymphoma in Malawi. *Pediatr Blood Cancer* 65:e27353. <https://doi.org/10.1002/psc.27353>
99. Chan SWW, Chien CW, Wong AYL et al (2021) Translation and psychometric validation of the traditional Chinese version of Patient-Reported Outcomes Measurement Information System Pediatric-25 Profile version 20 (PROMIS-25) in Chinese Children with Cancer in Hong Kong. *Qual Life Res*. <https://doi.org/10.1007/s11136-021-02759-8>
100. Yeh C-H, Hung L-C (2003) Construct validity of newly developed quality of life assessment instrument for child and adolescent cancer patients in Taiwan. *Psychooncology* 12:345–356. <https://doi.org/10.1002/pon.647>
101. Yeh C-H (2001) Adaptation in children with cancer: research with Roy's model. *Nurs Sci Q* 14:141–148
102. Verrips EGH, Vogels T, Koopman HM et al (1999) Measuring health-related quality of life in a child population. *Eur J Public Health* 9:188–193
103. Reeve BB, Hays RD, Bjorner JB et al (2007) Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 45:S22–31. <https://doi.org/10.1097/01.mlr.0000250483.85507.04>
104. DeWalt DA, Gross HE, Gipson DS et al (2015) PROMIS[®] pediatric self-report scales distinguish subgroups of children within and across six common pediatric chronic health conditions. *Qual Life Res* 24:2195–2208. <https://doi.org/10.1007/s11136-015-0953-3>
105. Jones CM, Baker JN, Keesey RM et al (2018) Importance ratings on patient-reported outcome items for survivorship care: comparison between pediatric cancer survivors, parents, and clinicians. *Qual Life Res* 27:1877–1884. <https://doi.org/10.1007/s11136-018-1854-z>
106. Forrest CB, Forrest KD, Clegg JL et al (2020) Establishing the content validity of PROMIS Pediatric pain interference, fatigue, sleep disturbance, and sleep-related impairment measures in children with chronic kidney disease and Crohn's disease. *J Patient Rep Outcomes* 4:11. <https://doi.org/10.1186/s41687-020-0178-2>
107. Garcia SF, Cella D, Clauser SB et al (2007) Standardizing patient-reported outcomes assessment in cancer clinical trials: a Patient-Reported Outcomes Measurement Information System initiative. *J Clin Oncol* 25:5106–5112. <https://doi.org/10.1200/JCO.2007.12.2341>
108. European Medicines Agency (2005) Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-regulatory-guidance-use-health-related-quality-life-hrql-measures-evaluation_en.pdf. Accessed 10 January 2023
109. Food and Drug Administration (2009) Guidance for industry: patient reported outcome measures: use in medical development to support labeling claims. <https://www.fda.gov/media/77832/download>. Accessed 10 January 2023
110. Rothman M, Burke L, Erickson P et al (2009) Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health* 12:1075–1083. <https://doi.org/10.1111/j.1524-4733.2009.00603.x>
111. Cocks K, Johnson C, Tolley C et al (2021) Evaluating content validity. <https://qol.eortc.org/projectqol/content-validity/>. Accessed 30 June 2021
112. Darlington AS, Sodergren S (2021) Adolescents and young adults. <https://qol.eortc.org/questionnaire/aya/>. Accessed 30 June 2021
113. Sireci SG (1998) The construct of content validity. *Soc Indic Res* 45:83–117
114. Edwards MC, Slagle A, Rubright JD et al (2018) Fit for purpose and modern validity theory in clinical outcomes assessment. *Qual Life Res* 27:1711–1720. <https://doi.org/10.1007/s11136-017-1644-z>
115. Lee E-H, Kang EH, Kang H-J (2020) Evaluation of studies on the measurement properties of self-reported instruments. *Asian Nurs Res (Korean Soc Nurs Sci)* 14:267–276. <https://doi.org/10.1016/j.anr.2020.11.004>
116. Ricci L, Lanfranchi J-B, Lemetayer F et al (2019) Qualitative methods used to generate questionnaire items: a systematic review. *Qual Health Res* 29:149–156. <https://doi.org/10.1177/1049732318783186>
117. Gagnier JJ, Lai J, Mokkink LB et al (2021) COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Qual Life Res*. <https://doi.org/10.1007/s11136-021-02822-4>
118. Terwee CB, Jansma EP, Riphagen II et al (2009) Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 18:1115–1123. <https://doi.org/10.1007/s11136-009-9528-5>
119. EORTC Quality of Life Group (2022) Development of an EORTC questionnaire for children with cancer (8–14 years). <https://qol.eortc.org/questionnaire/development-of-an-eortc-questionnaire-for-children-with-cancer-8-14-years/>. Accessed 04 Feb 2022

Publisher's Note

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