



ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

ISPOR TASK FORCE REPORTS

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

Louis S. Matza, PhD^{1,*}, Donald L. Patrick, PhD, MSPH², Anne W. Riley, PhD, MS³, John J. Alexander, MD, MPH⁴, Luis Rajmil, MD, PhD, MPH⁵, Andreas M. Pleil, PhD⁶, Monika Bullinger, PhD⁷

¹Outcomes Research, United BioSource Corporation, Bethesda, MD, USA; ²Department of Health Services, University of Washington, Seattle, WA, USA; ³Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; ⁴Division of Anti-Infective Products, FDA, Silver Spring, MD, USA; ⁵IMIM-Hospital del Mar Medical Research Institute, and Catalan Agency for Health Information, Assessment and Quality, Barcelona, Spain; ⁶Pfizer Global Pharmaceuticals, Pfizer, Inc., San Diego, CA, USA; ⁷Department of Medical Psychology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany

A B S T R A C T

Background: Patient-reported outcome (PRO) instruments for children and adolescents are often included in clinical trials with the intention of collecting data to support claims in a medical product label. **Objective:** The purpose of the current task force report is to recommend good practices for pediatric PRO research that is conducted to inform regulatory decision making and support claims made in medical product labeling. The recommendations are based on the consensus of an interdisciplinary group of researchers who were assembled for a task force associated with the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). In those areas in which supporting evidence is limited or in which general principles may not apply to every situation, this task force report identifies factors to consider when making decisions about the design and use of pediatric PRO instruments, while highlighting issues that require further research. **Good Research Practices:** Five good research practices are discussed: 1) Consider developmental differences and determine age-based criteria for PRO administration: Four age groups are discussed on the basis of previous research (<5 years old, 5–7 years, 8–11 years, and 12–18 years). These age groups are recommended as a starting point when making decisions, but they will not fit all PRO instruments or the developmental stage of every child. Specific age ranges should be determined individually for each population and PRO instrument. 2) Establish content validity of pediatric PRO instruments: This section discusses the advantages of

using children as content experts, as well as strategies for concept elicitation and cognitive interviews with children. 3) Determine whether an informant-reported outcome instrument is necessary: The distinction between two types of informant-reported measures (proxy vs. observational) is discussed, and recommendations are provided. 4) Ensure that the instrument is designed and formatted appropriately for the target age group. Factors to consider include health-related vocabulary, reading level, response scales, recall period, length of instrument, pictorial representations, formatting details, administration approaches, and electronic data collection (ePRO). 5) Consider cross-cultural issues. **Conclusions:** Additional research is needed to provide methodological guidance for future studies, especially for studies involving young children and parents' observational reports. As PRO data are increasingly used to support pediatric labeling claims, there will be more information regarding the standards by which these instruments will be judged. The use of PRO instruments in clinical trials and regulatory submissions will help ensure that children's experience of disease and treatment are accurately represented and considered in regulatory decisions. **Keywords:** adolescents, children, ISPOR, medical product labeling, patient-reported outcomes, pediatrics, PRO, task force.

Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

* Address correspondence to: Louis S. Matza, United BioSource Corporation, 7101 Wisconsin Avenue, Suite 600, Bethesda, MD 20814, USA.

E-mail: louis.matza@unitedbiosource.com.

1098-3015/\$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

<http://dx.doi.org/10.1016/j.jval.2013.04.004>

Background to the Task Force

In March 2009, the ISPOR Health Science Policy Council recommended to the ISPOR Board of Directors that an ISPOR Good Research Practices Patient-Reported Outcomes (PRO) Task Force should be established to focus on the Assessment of Patient-Reported Outcomes in Children and Adolescents. The Board of Directors approved this PRO Task Force in March 2009. The Pediatric PRO Task Force chair (Dr. Matza) and co-chair (Dr. Patrick) chose task force members based on their experience in PRO assessment and research focusing specifically on children and adolescents. Members were selected to represent a diverse range of perspectives, including government (United States Food and Drug Administration), academia, research organizations, and the pharmaceutical industry. In addition, the task force had international representation with members from Germany, Spain, and the United States.

The Task Force initially met approximately every 2 months by teleconference to develop an outline and discuss issues to be included in the report. Face-to-face meetings were held in New Orleans in October 2009 and Atlanta in May 2010 to discuss these issues further and come to consensus on recommendations. In addition, the task force chair had a series of one-on-one teleconferences with members involved in drafting the manuscript. All task force members reviewed many drafts of the report and provided frequent feedback in both oral and written comments.

Preliminary findings and recommendations were presented in a forum at the ISPOR 15th Annual International Meeting in May 2010. Updated findings and recommendations were presented in a forum at the ISPOR 17th Annual International

Meeting in June 2012. Comments received during these two forums were addressed in subsequent drafts of the report.

A draft of this report was distributed to the ISPOR PRO Review Group (which includes over 400 members) in March 2012. A revised draft was distributed to the entire ISPOR membership in January 2013. During these two rounds of review, over 250 written comments were received from 40 ISPOR members. Written comments were also provided by members of several regulatory and reimbursement agencies including three reviewers from the US Food and Drug Administration, one reviewer from Germany's Institute for Quality and Efficiency in Health Care (IQWiG), and one reviewer representing both the French National Authority for Health (Haute Autorité de Santé [HAS]) and the European Network for Health Technology Assessment (EUnetHTA).

All comments were considered, and most were substantive and constructive. The comments were discussed by the task force in a series of teleconferences and addressed as appropriate in revised drafts of the report. Once consensus was reached by all task force members, the final report was submitted to *Value in Health* in April 2013.

All written comments are published at the ISPOR Web site on the task force's Web page: <http://www.ispor.org/TaskForces/PRO/ChildrenAdolescents.asp>. The task force report and Web page may also be accessed from the ISPOR homepage (www.ispor.org) via the purple Research Tools menu, Good Practices for Outcomes Research, heading: Patient Centered & Clinician Reported Outcomes Methods, and link: Assessment of PRO in Children and Adolescents. A list of reviewers is also available via the task force's Web page.

Introduction

A patient-reported outcome (PRO) instrument involves the report of health status coming directly from the patient without interpretation of the patient's response by a clinician, investigator, or anyone else [1,2]. Many aspects of medical conditions are known only by the patients themselves, and direct assessment of the patient perspective is necessary to thoroughly understand patients' experiences of disease and treatment. In recent years, there has been an increased emphasis on systematic development and validation of PRO instruments for use in clinical trials evaluating medical product efficacy. PRO instruments are often included in clinical trials with the intention of collecting data to support claims made about a medical product in the product label [3,4].

The US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have released guidelines for the assessment of PROs. The FDA guidance has had a strong influence on industry-funded PRO research since a draft was published in 2006 and finalized in 2009 [1]. This guidance provides an overview of PRO use in the context of medical product development, as well as guidance for developing and evaluating these instruments. A brief section of the guidance discusses PRO instruments intended for use with children and adolescents (Section III.G.1). This section begins by stating that "issues related to the development process for pediatric PRO instruments are similar to the issues detailed for adults." Then, the section continues by saying that the use of PRO instruments in pediatric populations introduces unique challenges that are not encountered in PRO research with adults. Several challenges are mentioned, including age-related vocabulary, comprehension of health concepts, the need to determine the lower age limit at which children can provide reliable and valid responses, and appropriate use of reports by informants other than the patients themselves. No specific recommendations, however, are provided for addressing these challenges.

Like the FDA, the EMA has provided recommendations for PRO measurement, particularly with regard to the assessment of health-related quality of life (HRQOL) [5]. The EMA guidance, however, does not discuss the use of PRO instruments with children and adolescents. In sum, there is limited available guidance for research involving pediatric PRO assessment related to medical product development.

Therefore, the purpose of the current task force report is to recommend good practices for pediatric PRO research that is conducted to inform regulatory decision making and support claims made in medical product labeling. The recommendations in this report are based on the consensus of an interdisciplinary group of researchers who were assembled for a task force associated with the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). The good research practices are summarized in Figure 1.

The challenges of choosing, developing, and implementing PRO instruments in children and adolescents have been reviewed and discussed in many previous publications [6–22]. Several published articles and book chapters have also provided lists and reviews of generic and condition-specific PRO instruments for children and adolescents [6,9,14,23–28]. The current task force report differs from this previous work because of its specific focus on pediatric PRO instruments in the context of medical product development and labeling.

The recommendations in this report are based on published research as much as possible. Pediatric PRO assessment, however, is a developing field of research, and empirical evidence is limited for some important areas of instrument design, development, validation, and implementation. Therefore, it is not currently possible to provide definitive recommendations for some of the issues discussed in the current report. In these situations, this task force report discusses the factors to consider when making decisions about the design and use of PRO instruments for children and adolescents. In addition, this report highlights areas in which further research is needed to advance the field of pediatric PRO assessment.

Good Research Practices	Comments and Recommendations
1. Consider Developmental Differences and Determine Age-Based Criteria for PRO Administration	<ul style="list-style-type: none"> • Four age groups are discussed. These age groups should be used as a starting point when making decisions. It is not possible to provide age cutoffs that will fit every situation. Specific age cutoffs should be determined individually for each PRO instrument and tested with cognitive interviews in each new target population. • Less than 5 years old: No clear evidence of reliability or validity of child-report measures • 5 to 7 years old: Child-report is possible, but reliability and validity are often questionable • 8 to 11 years old: Reliability and validity of child-report improves • 12 to 18 years old: Self-report is preferred
2. Establish Content Validity of Pediatric PRO Instruments	<ul style="list-style-type: none"> • Children and adolescents can be effective content experts. • In most cases, children should be included in qualitative research performed to establish content validity of pediatric PROs. • Cognitive interviews should be conducted with the intended respondent. Children should be interviewed for child-report instruments, and parents should be interviewed for parent-report instruments. • Content validity should be demonstrated within narrow age groupings.
3. Determine Whether an Informant-Reported Outcome Instrument is Necessary	<ul style="list-style-type: none"> • Informant-reported outcomes include both proxy and observational measures. • When children in the target age range are capable of completing a PRO instrument independently, a child-reported measure should be used. • Second, when children in the target age range are not capable of completing a PRO measure, an informant-reported measure may be used. • Informant-reported measures should assess observable content as much as possible.
4. Ensure that the Instrument is Designed and Formatted Appropriately for the Target Age Group	<ul style="list-style-type: none"> • Health-related vocabulary and reading level • Response scale • Recall period • Length of instrument • Pictorial representations • Formatting • Administration approaches • Electronic data collection (ePRO)
5. Consider Cross-Cultural Issues	<ul style="list-style-type: none"> • Content validity and measurement approach of a pediatric PRO instrument will need to be re-examined within each new culture.

Fig. 1 – Good research practices discussed in this Task Force Report.
PRO, patient-reported outcome.

Good Research Practices in Developing and Implementing Pediatric PRO Instruments

Good Research Practice 1: Consider Developmental Differences and Determine Age-Based Criteria for PRO Administration

The pediatric PRO literature has frequently focused on questions of developmental differences and age-related cutoffs

[16,21,29,30]. For example, at what age can children begin to report their health status, at what age are their responses reliable, and when can children respond to items addressing more abstract constructs? This task force group does not think that it is possible to provide age cutoffs that will apply in all situations. For any individual study or PRO measure, the answer to these age-related questions will depend on a range of factors, including the concept being assessed, the format of the instrument, the level of independence required of child respondents

(e.g., self-reported questionnaires vs. interviewer-administered measures), and characteristics of the population of interest (e.g., medical or psychiatric condition). Determination of age cutoffs is further complicated by the substantial cognitive and developmental variability within any particular age group, as well as possible gender differences in perceptions and management of medical conditions [31].

Despite these challenges, it is possible to offer guidance regarding the abilities of children in various age ranges on the basis of research in the areas of psychometrics and cognitive development. While no age cutoffs will be appropriate for all PRO instruments, it is necessary to discuss general age ranges to summarize the literature and provide practical recommendations. This task force proposes that pediatric samples can be categorized into four general age groups with regard to PRO assessment for use in medical product evaluation, but these age groups should be used only as a starting point when making decisions. Specific age cutoffs should be determined individually for each PRO instrument, and age ranges for a PRO instrument should be tested with cognitive interviews whenever the instrument is applied in a new target population.

Furthermore, as stated above, children within each age group are likely to vary in their ability to reliably complete questionnaires. Researchers will need to remain aware of this variability during data collection. While some studies rely on the judgment of interviewers or data collection staff to determine whether individual children can complete the study, a more standardized approach is to screen for children's ability to complete PRO measures. For example, one recent study assessed the reading level of every child with the American Guidance Service (AGS) Reading Level Indicator [32]. Children who demonstrated reading ability at or above the fourth-grade level were offered the option of completing the study measure with a written mode of administration. For children scoring below the fourth-grade reading level, the study measure was administered by an interviewer. This screening approach is based on the idea that each child's ability, rather than numerical age, should determine if and how a PRO measure should be administered. This task force encourages further research on ways to screen for PRO instrument completion skills and address the variability within age groups.

It should also be noted that there is no clear consensus on definitions of childhood or adolescence. Definitions vary by culture and governmental organization. The definition adopted by the World Health Organization [33] was originally provided by the United Nations' Convention on the Rights of the Child, Article 1: "A child means every human being below the age of 18" [34]. The Centers for Disease Control and Prevention has cited this definition at least once [35], but a different Centers for Disease Control and Prevention document has focused on a broader group including individuals up to 19 years old [36]. The older part of this range, beginning at approximately age 12 or 13 years, is usually labeled as "adolescence" [37,38]. While acknowledging that these definitions are fluid and variable, this task force is focused primarily on the age range of 0 to 18 years. In the current report, the term "adolescence" refers to individuals ranging from roughly 12 to 18 years old, while the term "younger children" generally refers to those aged below 8 years.

Age group 1 (younger than 5 years old)

The first age category begins at birth and includes infants, toddlers, and preschoolers. Although self-report instruments have been designed to be completed by children younger than 5 years, there is no clear evidence of reliability or validity of self-report measures in this age group [39]. Therefore, the assessment of health status in this youngest age range must rely on clinical measures and observational reports of parents or other adults.

Age group 2 (5–7 years old)

The second age category consists of younger children ranging from age 5 to 7 years. Generic and condition-specific self-report measures are available for assessing the health status, including symptoms and HRQOL, of children in this age range [9,10,14,24,27,28,40]. Some researchers have advocated for direct assessment of the child's perspective beginning in this age range [8,30,41], and there are some reports of acceptable psychometric data as low as age 5 years. For example, 5-year-old children completing the Pediatric Quality of Life Questionnaire (PedsQL), a widely used generic HRQOL measure, yielded minimal missing data (2.8%) and good internal consistency reliability with a Cronbach's alpha of 0.86 for the total score based on all 23 items [21]. Results of other studies, however, have contradicted this finding, with reports of less encouraging psychometric results. In the validation of the child-report version of the Child Health and Illness Profile (CHIP-CE), internal consistency and test-retest reliability for children aged 6 to 7 years were below a generally accepted range (i.e., ≥ 0.70) for three of five subscales (Cronbach's alphas for multiple subscales ranged from 0.64 to 0.83; intraclass correlation coefficients [ICCs] = 0.35–0.69), but these reliability statistics were notably better in the 8 to 11 years age group (Cronbach's alphas = 0.72–0.85; ICCs = 0.71–0.80) [42]. Similarly, developers of a self-report measure for children with epilepsy reported an increase in test-retest reliability from children aged 6 to 7 years (ICCs = 0.18–0.52) to children aged 8 to 15 years (ICCs = 0.59–0.69) [43]. Furthermore, a review of 53 health-related self-report measures for children up to age 8 years found that only 51% of the measures met minimum standards for internal consistency reliability (i.e., Cronbach's alpha = 0.70), and only 23% of the measures met standards for test-retest reliability (i.e., ICC = 0.70) [24]. Based on available data from young children, test-retest reliability seems to be particularly problematic, whereas internal consistency reliability tends to be closer to generally accepted standards.

In addition to concerns about reliability, some findings suggest that many children aged 5 to 6 years may not use the full range of a typical five-point Likert scale. Instead, they may rely more on responses at the middle and/or extreme ends of the scale, as if they were responding on a three-point scale [19,44]. Health-related vocabulary may also be limited in children of this age range. A cognitive interview study found that children's comprehension of terms commonly used in health status questionnaires (e.g., nervous, energy, and pain) tends to be limited at age 5 years, but gradually increases from ages 6 to 8 years (Fig. 2) [19].

Research on cognitive processing capabilities suggests that PRO data reported by children in this age range may be useful for assessment of some concepts but should be interpreted with caution [29,45]. Children aged 5 to 7 years typically understand basic ideas of feeling healthy/unhealthy or good/bad, and they can usually apply these ideas to themselves, often with some specificity. It has also been suggested, however, that children in this age range may have uneven cognitive capacities (i.e., strong abilities in one cognitive domain such as word recognition, but not in another area such as attention). Furthermore, some children may be slow, inefficient, or error-prone in choosing the response option that best represents their perspective [29,45].

In sum, although self-report measures may be useful for gathering health-related information among children aged 5 to 7 years, previous research raises questions regarding the reliability and accuracy of their responses. For PRO measures to be considered acceptable for regulatory decision making, they will need to demonstrate content validity and acceptable psychometric properties in the target age range, which may be challenging in this age group. The likelihood of demonstrating acceptable measurement properties in children aged 5 to 7 years may be maximized by using or designing PRO measures that assess simple concepts

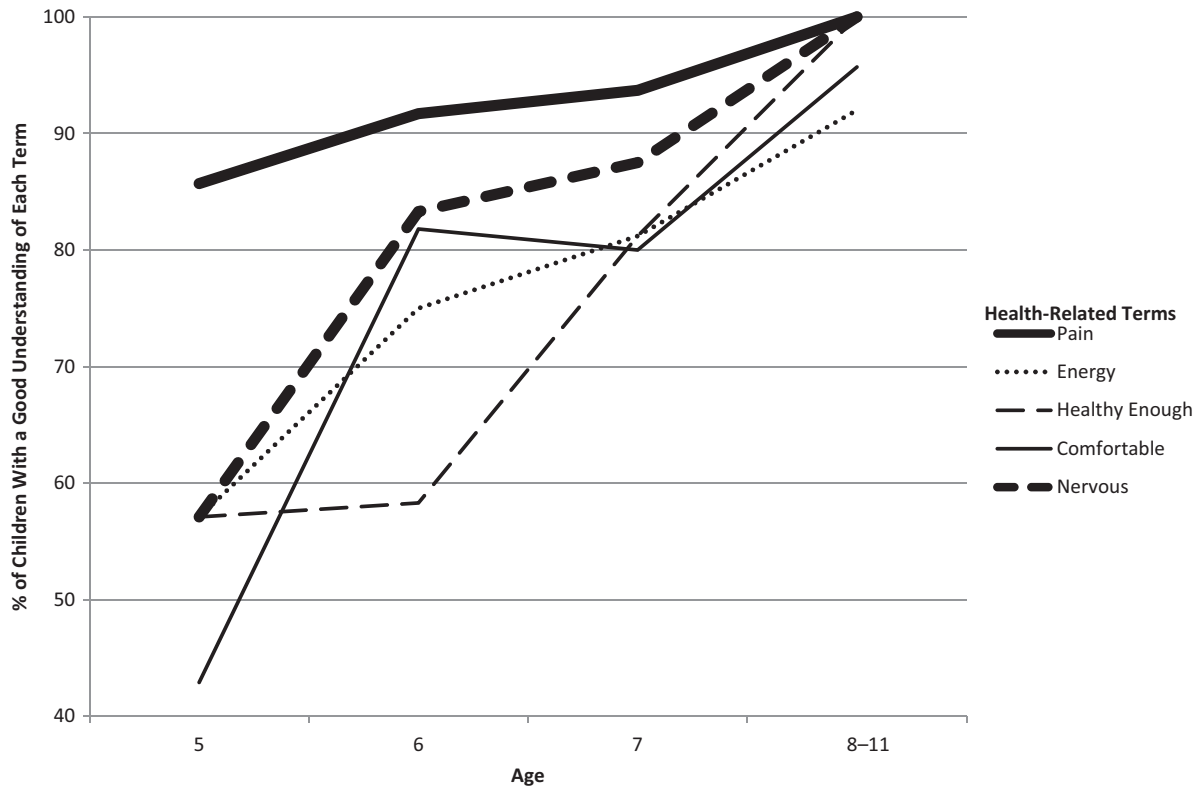


Fig. 2 – Age differences in children’s understanding of health-related terms.

Findings reported by Qual Life Res, 10, Rebok G, Riley A, Forrest C, et al., Elementary school-aged children’s reports of their health: a cognitive interviewing study, 59–70, 2001. With kind permission from Springer Science and Business Media.

with clear formatting, age-appropriate vocabulary, and fewer response options than do measures for older children. If it is not possible to demonstrate adequate measurement properties in this age group, informant-reported outcome measures completed by parents or other adults may be used (see Good Research Practice 3 of this task force report). We encourage further research on child-reported PRO measures in this age group in the hope that children’s experience and subjective perspective may be considered in regulatory decision making.

Age group 3 (8–11 years old)

The third age group consists of older children, ranging from roughly age 8 years to age 11 years. Children’s ability to independently complete health status measures appears to improve in this age range. As described earlier in the discussion of age group 2, several studies have found increased reliability, comprehension of health vocabulary, and understanding of Likert scales in samples of children aged 8 years and older than in samples of younger children [19,42,43]. Self-reports of children in this older age range frequently meet generally accepted standards for reliability on both generic and condition-specific measures [21,27,46].

Results of a recent cognitive interview study conducted by the Patient Reported Outcomes Measurement Information System (PROMIS) pediatrics group demonstrate some of the relevant skills of children in this age group [47]. The sample in this qualitative study focusing on a pediatric item bank included 24 children aged 8 or 9 years. Like the adolescents, children in this age group were able to comprehend the majority of items, response options, directions, and recall period, while identifying language that was difficult to understand.

Research on children’s cognitive development provides further support for the use of self-report methodology for children beginning in the 8 to 11 year age range [29]. At this point of development, the cognitive capacities needed to respond to health questionnaires have developed in most children. These capacities include 1) an understanding of health concepts; 2) self-regulatory abilities including sustaining effort, minimizing frustration, and avoiding distractions; and 3) cognitive processing capacities including the ability to understand the task, keep the question in working memory long enough to respond, evaluate how the question applies to oneself, evaluate the differences between response options, and choose the optimal response [29].

In sum, there is a growing body of research supporting the feasibility and reliability of self-reports in children aged 8 to 11 years. Therefore, it may be acceptable to use self-report data from this age group in research conducted for medical product evaluation. As with all PRO end points, careful qualitative research will be required to convincingly demonstrate that children in this age range understand the PRO measure as intended.

Age group 4 (12–18 years old)

The fourth age group consists of adolescents, beginning at approximately age 12 years. Most adolescent PROs are designed and validated for respondents up to ages 16, 17, or 18 years [27]. Adolescent self-report measures do not appear to face greater challenges meeting generally accepted psychometric standards than do adult measures [21,28,46]. The primary challenge in developing PRO measures for adolescents, however, is capturing content that is relevant to this age group. Although items assessing symptoms may be similar to those in adult measures, developers designing instruments to assess symptom impact or

functional status will need to consider the unique social and emotional aspects of this developmental stage. This was illustrated by a qualitative study that used focus groups and interviews with 12- to 18-year-olds to elicit concepts for a quality-of-life measure [48]. The resulting conceptual model highlights the uniqueness of adolescence, with a wide range of relevant concepts such as adult support, freedom, peer relations, being oneself, spirituality, life satisfaction, monetary resources, view of the future, and education. Many of these domains are more sophisticated than concepts that may be included in a measure aimed at children below age 12 years. Yet, domains such as peer relations, view of the future, and education may include content different from that of adult PRO measures. Thus, many adult measures are unlikely to be appropriate for adolescents. Determining appropriate item content for this age group is further complicated by the heterogeneity across this age span. For example, content that is relevant to an 18-year-old may not be relevant to a 12-year-old. In sum, adolescent self-report measures frequently demonstrate the psychometric properties necessary to be used as a key study end point in clinical trials. Qualitative research, however, will also be required to demonstrate that an individual instrument has content validity specific to the target age group.

Individuals older than 18 years are generally administered PROs that were designed for use across all ages of adult samples. However, it may be useful to consider the relevance of previously validated instruments to young adults who may have distinct developmental characteristics. Researchers have suggested that the 20s could be viewed as a stage of life called “emerging adulthood” [49,50]. This phase of life appears to be associated with unique challenges with regard to the level of independence, personal identity, career development, and relationships. Research on development and adaptation during this phase will be published in a new scientific journal titled *Emerging Adulthood*, which is launching in 2013. Some adult PRO measures may require qualitative research to ensure that all items have content validity in this transitional age group.

In sum, when developing new PRO instruments or using existing instruments for research involving medical product evaluation, it is critical to demonstrate the age appropriateness of the instrument for the target population. Age appropriateness of a PRO should be documented with a combination of qualitative and quantitative research. First, cognitive interviews should be conducted to examine that all aspects of the instrument are appropriate for the target age group. These interviews should assess the respondents’ perspectives on the instrument’s vocabulary level, item content, recall period, response options, instructions, comprehensiveness, relevance, and clarity. This qualitative research should be conducted with an adequate sample size at the upper and lower bounds of the target age range to ensure that there are sufficient data to support the selected age cutoffs. Subsequently, quantitative research can be conducted to examine the instrument’s psychometric properties in a sample matching the intended age range of children in planned clinical trials.

Good Research Practice 2: Establish Content Validity of Pediatric PRO Instruments

Children as content experts

In recent years, establishing content validity has been emphasized as a critical step of PRO instrument development and validation [1,51,52]. Definitions of content validity vary [53], but the definition proposed by a previous ISPOR task force appears consistent with most of the literature, including the FDA guidance: “The extent to which an instrument contains the relevant and important aspects of the concept it intends to measure” [54]. There is also a consensus that content validity is primarily

established through qualitative research that includes direct input from the target population.

This qualitative work, which has previously been described in detail, generally proceeds in two sequential steps: concept elicitation and cognitive interviews [1,55,56]. First, concept elicitation interviews or focus groups with individuals from the target patient population are conducted to inform item generation [51]. Semi-structured interview and focus group discussion guides are usually drafted on the basis of information gathered during literature review and clinician interviews. In the concept elicitation phase, patients help identify concepts and wording that will shape the items of a PRO measure. When conducting concept elicitation work for pediatric PRO measures, it is particularly important to identify the specific language patients use to describe their illness and treatment because children and adolescents may use different words than do adults. Information gathered during concept elicitation should also be used to support the instrument’s conceptual framework, which is a diagram illustrating relationships between items, domains, and concepts measured by the instrument [1]. It is important that the conceptual framework of a pediatric instrument be specifically relevant to the target age group, rather than a replication or rewording of a conceptual framework previously developed for an adult measure [14].

After drafting an instrument, the second step is to conduct cognitive interviews to assess and refine the draft instrument on the basis of patients’ perceptions of the measure’s relevance, clarity, and comprehensiveness [52,57]. These qualitative research procedures are considered necessary for establishing content validity, which cannot be demonstrated via quantitative psychometric analyses. The FDA guidance provides detailed recommendations regarding methods for establishing content validity, but it does not mention conceptual or methodological challenges unique to the assessment of content validity for pediatric measures.

Despite the FDA’s encouragement to provide documentation of content validity based on patient input [1], researchers may still be hesitant to develop PRO instruments on the basis of direct input from children. Content validity of child- and parent-reported pediatric measures has often been supported by input from parents or clinical experts rather than the children themselves [58–61]. A growing body of research, however, has indicated that children and adolescents can be effective content experts [62–65]. For example, one study examined the use of adolescents as “experiential experts” when establishing content validity, and the authors concluded that the adolescents’ input resulted in greater relevance of the PRO for the target population [66]. Clearly, there are situations when it is not feasible to obtain the child’s perspective, such as when the target population is too young or cognitively impaired to read, comprehend, or respond to questions about health status [67,68]. When children, however, are able to discuss their health status, they may have a unique perspective based on personal experience that is unknown to clinicians or parents.

As with adult PROs, assessment of the child’s perspective during concept elicitation and item generation is critical for ensuring that a pediatric PRO measure is relevant and comprehensive to the target population. Furthermore, there are some aspects of symptom experience and impact that are known only to patients themselves [2]. Therefore, we recommend either including children in qualitative research performed to establish content validity of pediatric PROs or providing justification for excluding children from this research. Qualitative data from children should be considered along with parent, clinician, and/or expert reports when drafting and refining measures.

The central role of children’s context

The basic principles and methodology of content validity are the same for pediatric PROs as for adult PROs. The assessment of

content validity for pediatric measures, however, is associated with unique conceptual issues and methodological challenges that do not arise when developing measures for adult populations. First, the early phases of establishing content validity focus on identifying the target concept of the PRO measure and the “content” of the items. When conducting research to elicit concepts and items for a pediatric measure, it is important to consider the central role of context in child development. Children are embedded within multiple social contexts that may be different from those of adults. These contexts include the family, the child’s peer group, the classroom, extracurricular activities, and the community [16,69]. The impact of disease and treatment, as well as the relevance of individual symptoms, may be different within each of these contexts. For example, there may be symptoms that have a relatively minor impact on adults, but limit a child’s ability to participate in athletics or daily activities with peers. Thus, when developing pediatric PRO instruments, researchers must consider the relevance and impact of symptoms and treatment within these contexts. Although parents may have insight into their children’s health experiences at home, only the children themselves are likely to have a broad perspective spanning multiple settings and social groups. Therefore, to ensure that an instrument adequately assesses all important aspects of a condition and its treatment, the child’s perspective should be considered when establishing content validity.

Concept elicitation and item generation

Procedures for concept elicitation and item generation have previously been described in detail [51]. Concept elicitation research with children, however, raises additional methodological issues that are not encountered when conducting qualitative research with adult samples. First, we recommend conducting this qualitative research with several types of respondents, including the children themselves. As previously stated, children should be included in the early stages of establishing content validity whenever possible, particularly when developing a child-report instrument. Clinicians, including pediatricians, pediatric nurses, and other medical staff, are another important source of information. When developing a pediatric PRO, it is necessary to query clinicians who have direct experience treating children in addition to experience with the relevant medical condition. Parents or other primary caregivers are also likely to be a critical source of information, particularly when developing measures for younger children or when developing parent-reported measures. Finally, when assessing conditions such as attention-deficit/hyperactivity disorder that may be particularly relevant in the school setting, teachers may have unique insight into children’s symptoms and the impact on social and academic functioning. In sum, content validity of pediatric PROs should be based on qualitative data from patients as well as other informants such as parents and clinicians.

Concept elicitation research may be conducted via individual interviews or focus groups, and these two approaches have different advantages [51]. For example, focus groups allow researchers to reach many patients at once, while patients can use the ideas of other respondents as cues to express their own views. With adult patients, focus groups are sometimes recommended so that participants can respond to each other, thus leading to richer information about the patient experience [56]. Some challenges inherent to the group process, however, may arise more with children or adolescents than with adults. For example, some younger children may become distracted to a greater extent than do older children or adults. Children’s reports may also be influenced by their peers’ statements or a social desirability bias, resulting in a less accurate presentation of their own symptoms and experience. Furthermore, because of

the complex social pressures of childhood and adolescence, pediatric respondents may be reluctant to speak openly in group discussion.

When conducting focus groups, researchers can take steps to encourage children and adolescents to share accurate and honest health-related information. For example, although many adult health issues can be discussed in mixed gender groups, children and adolescents may be more comfortable and forthcoming in single gender groups, particularly when discussing sensitive topics such as sexual activity or adolescent maturation issues. Researchers should also consider whether the medical condition of interest may cause shame, embarrassment, or self-consciousness, which could be more inhibiting for children and adolescents than for adults. For example, adolescents may resist discussing their difficulties with obesity in the presence of peers, and one qualitative study reported that children with foot and ankle problems expressed self-consciousness regarding the appearance of their shoes, which were different from those of other children [62]. Adults may also be hesitant to discuss some topics in the presence of other patients, but this hesitance or embarrassment can be particularly challenging with children and adolescents.

Although individual interviews do not benefit from the group dynamic of focus groups, interviews may be used to gather more detailed information about an individual’s experience. Interviews may be more effective than focus groups for sensitive topics that may be uncomfortable to discuss in front of others. In addition, unlike focus group results, data from individual interviews can be used to accurately represent each individual respondent when creating a saturation table or grid to support content validity (see previous ISPOR task force report for details regarding saturation tables; Patrick et al. [51]). In contrast, when analyzing focus group data, it is often difficult to determine precise numbers of patients who reported a specific symptom or impact.

This task force recommends that child concept elicitation interviews conducted to inform the development of child-reported PROs should be performed without parents whenever possible. The purpose of these interviews is to understand the child’s own subjective perspective and experience, and parent participation in the child interview session could introduce a risk of obscuring the child’s perspective. For example, some parents may impose their own perspective or be tempted to answer questions on behalf of their children, rather than letting children speak for themselves. Furthermore, if children know a parent is listening, they may feel inhibited when discussing certain topics, or they may be motivated to please their parents rather than responding honestly. For most child-reported PRO measures, the parent’s perspective is likely to be useful for eliciting concepts and generating items. This task force recommends that parents be interviewed separately from children to assess the parents’ views without having an impact on information gathered during the child interview.

There are some context-specific exceptions to the recommendation that child concept elicitation interviews be conducted without a parent present. If a child is not comfortable being interviewed without a parent present, researchers should allow the parent to join the child in the interview room, but the parent should be discouraged from actively participating in the interview process. Parents may also be asked to sit behind the child so that the parent cannot make eye contact with the child or inadvertently influence the child’s responses. In addition, there are rare situations when a questionnaire is designed to be administered to a parent-child dyad to assess a shared medical experience [70]. In these situations when the shared perspective is of interest, it may be useful to develop the questionnaire based on dyadic rather than individual concept elicitation interviews. It should also be noted that there is some support for a parent-child dyad approach to assessment of the child’s HRQOL, as opposed to

the concept elicitation phase of PRO instrument development. Some innovative studies have suggested that trained interviewers can elicit reliable and valid responses from parent-child dyads, while possibly enhancing accuracy and minimizing response bias and missing data that can occur when children complete measures or interviews independently [71,72]. While these thought-provoking findings clearly merit consideration and future research, this task force believes that the risks of parent participation in children's concept elicitation interviews outweigh the potential benefits except in rare circumstances. Therefore, it is recommended that children's concept elicitation interviews be conducted without parent participation (or participation of other adults such as teachers) whenever possible to ensure that the child's responses are not biased or inhibited by the parent's opinions and presence.

Given the challenges of conducting focus groups with children and adolescents, researchers may choose to conduct both focus groups and individual interviews. This combined approach takes advantage of the rich information that may arise from the group dynamic in a focus group, while confirming that individual interviews do not yield different concepts. If focus groups appear to be unproductive for specific medical conditions within certain age groups, researchers may need to continue qualitative research using only individual interviews.

Designing a productive concept elicitation interview or focus group for children is often challenging. At the beginning of interviews/focus groups, the interviewer should speak with the children to confirm that they understand the purpose of the interview. In addition, direct inquiries regarding symptoms and their impact can be productive only if they are worded and presented in a way children can understand. One approach is to begin with general questions and gradually proceed to more focused queries. For example, initial questions may be most effective if they are simple, open-ended, and easy to answer, while conveying an interest in the child's life (e.g., "Can you tell me about what you do at home after school?"). Follow-up questions could then assess the impact of health and illness on activities and other topics raised by the child in response to the introductory questions. Subsequent questions can gradually focus more specifically on the child's health and related limitations. During the concept elicitation process, researchers should remain open to improvements in the interview design, and the interview guide may need to be revised on the basis of the first few interviews.

Some researchers have suggested creative activities to help children provide useful information during the interviews or focus groups. One interesting example is the focus group study conducted to elicit concepts for the Oxford Ankle Foot Questionnaire, which assesses disability associated with children's foot and ankle problems [62]. In one part of these groups, children participated in a "life-mapping" activity in which they were asked to consider a day in the life of a hypothetical child with a foot or ankle problem, including issues related to various times of day and multiple settings (e.g., in the morning, at school, at home, weekends, and holidays). Children were asked to name this hypothetical child and work together to describe problems that this child encountered in each setting. The researchers asked the children to discuss a hypothetical child to help them discuss potentially sensitive issues, such as bullying, without directly asking about difficult personal experiences. As another strategy for making the children feel comfortable, these focus groups began with a warm-up activity involving discussion of a non-threatening topic. This activity appeared to have beneficial effects on children's willingness to speak freely in the group. The authors concluded that focus groups involving creative child-centered activities elicited useful information regarding the effects of childhood foot and ankle problems, even with most of the children in the youngest age group (5–7 years old).

Another qualitative study used an innovative approach involving drawing to elicit descriptions of restless legs syndrome and its impact in children aged 6 to 17 years [73]. Children were asked, "Do you think you could draw me a picture of how it feels when you get that feeling in your legs on this piece of paper?" After completing the drawing, children were requested to "Tell me about your picture." The drawings appeared to facilitate communication and help the children describe their symptoms verbally. These results suggest that drawing could be a potentially useful approach in concept elicitation interviews or focus groups with children.

Cognitive interviews

After drafting a PRO instrument, it is standard practice to conduct interviews with respondents from the target population to provide additional evidence of content validity by confirming that patients understand the items and believe the measure adequately represents the concept of interest [1,52,54]. These interviews have frequently been called "cognitive debriefing interviews," but they are more recently referred to as "cognitive interviews." For pediatric instruments, we recommend that cognitive interviews be conducted with the intended respondent. That is, children should be interviewed regarding child-report instruments, and parents should be interviewed regarding parent proxy or observational instruments.

In cognitive interviews for adult instruments, respondents in the target population are asked about the clarity, comprehensibility, comprehensiveness, and relevance of items. Cognitive interviews with children can follow the same general interview procedures as those previously described for adults [52,55,56,74], but it is also useful to probe further to ensure that the interviewer truly understands the child's level of comprehension. For example, after a child has responded to an item, the child may be asked to explain how he or she selected the answer. Other useful questions about individual items may include "what does this question mean to you?" and "what is this question asking you to do?" The child's responses may reveal that the child had interpreted the question differently than the researcher had initially thought. A useful approach for characterizing the types of difficulties respondents may have with items has been presented by Knafel et al. [74].

The developmental appropriateness of a PRO instrument is a central issue to be considered during the cognitive interview process. The FDA guidance [1] recommends "instrument development within fairly narrow age groupings." This recommendation is particularly important for cognitive interviews. Even if a PRO instrument has demonstrated content validity in one age range, content validity in other age ranges cannot be assumed. Although the content needed to represent a given construct may be similar across age groups, developmental differences in the understanding of particular items and in the relevance of each item may affect content validity. For example, vocabulary that is commonly used by children of one age group to describe their symptoms may be too sophisticated for slightly younger children. Furthermore, the content of some items may not be equally relevant across all child age groups, particularly items assessing social or activity-related impact of symptoms. A previous ISPOR task force has provided detailed recommendations on cognitive interviewing methods for evaluating other aspects of PRO instruments such as instructions, recall period, response options, instrument formatting, instrument length, and mode of administration [52]. When conducting these interviews for child-reported instruments, it is important to evaluate each of these characteristics within narrow age groupings to ensure that all aspects of the PRO measure are appropriate for the full age range of the intended population. In sum, whether developing a new instrument or using a preexisting instrument

with pediatric samples, cognitive interviews must document content validity, clarity, comprehensibility, and age appropriateness of a PRO instrument throughout the specific age range matching the intended use of the instrument.

Throughout a cognitive interview, children may have more difficulty than adults in understanding their role as content experts in the questionnaire development process. For example, some children may focus on their own responses, rather than the more abstract task of reviewing the questionnaire for content and clarity [64]. Thus, clear explanation of the questionnaire review process should be provided at the beginning of each interview, and interviewers should be trained to reorient the child to the task of reviewing the questionnaire whenever necessary.

Although there are few published examples of cognitive interview studies conducted with children, some studies have reported encouraging results. In the process of developing a pediatric item bank, the PROMIS pediatrics group conducted cognitive interviews with 77 children ranging from ages 8 to 17 years [47]. Children in all age ranges were able to contribute useful information regarding their own health and the items themselves. The authors also concluded that children offered a perspective on items' comprehensibility, as the younger children helped to identify several terms that were difficult to understand. A cognitive interview study of the PedsQL Diabetes Module also included interviews with children as young as 8 years [75]. The version of the questionnaire for 8- to 12-year-old children was edited on the basis of children's input during these cognitive interviews.

Good Research Practice 3: Determine Whether an Informant-Reported Outcome Instrument Is Necessary

When assessing younger children's health status, it is common to assess the perspective of informants instead of the children themselves. Informant-reported outcome measures may be necessary when children are not able to complete PRO measures reliably on their own because of their developmental stage, illness severity, language ability, or cognitive functioning [16,29,76,77]. This task force uses the term *informant* to refer to people other than the child who provide information related to the child. The informant is most frequently the child's parent, but measures may also be completed by other adults such as clinicians, teachers, and other caregivers. Although measures completed by informants are not actually "patient-reported," this task force provides recommendations for the use of informant measures because of their frequent use in pediatric assessment.

Before providing recommendations, it is necessary to define some relevant terminology. Informant-reported measures can be divided into two broad categories on the basis of the content of the items in the method of assessment: *proxy* and *observational*. These two categories can be further divided into subcategories, as illustrated in Figure 3. Proxy items require the informant to make inferences about the child's subjective experience. Many parent-report forms include proxy items asking parents to report their child's experiences of subjective constructs, such as emotional state, level of satisfaction, or pain severity [78–80]. Proxy items may also ask an informant to report impressions of how a child would answer specific questions. One study focusing on caregivers of adult patients found that items requiring informants to respond as if they were the patient may have a stronger correlation with patient reports than may items asking informants about patients' subjective experience [81]. In pediatric research, proxy items may yield useful data as parents typically have a valuable understanding of their children based on many important experiences. Furthermore, the parents' perspective is important because parents typically make health care decisions for children and adolescents.

The literature on proxy reporting has often focused on the degree of agreement between parent and child reports. This research has found mixed results, with some studies reporting strong agreement [82–85], while other studies have highlighted discrepancies between parent and child perspectives [86–92]. The degree of agreement appears to depend on a complex combination of factors, including the parent's health [86,91,93], parental distress [94], the child's age [21,86,95], the child's health [25,90,93,96], the statistical method for assessing agreement [86,93], and the domain being assessed [25,93]. Given these findings, most researchers recognize that parent reports cannot be accepted as a consistently accurate representation of children's subjective health experiences. Although parents can provide a useful perspective on their children's health status, proxy-report items inherently require the respondent to make inferences about another person's subjective experience, and these inferences are not necessarily equivalent to the person's actual experience.

Because of these challenges, the FDA PRO guidance [1] discourages proxy-reported outcome measures for pediatric populations in the context of research on medical product evaluation. The guidance, however, acknowledges that there are pediatric populations, such as very young children, who will not be able to report their own health status. In these situations when it is necessary to gather information from informants, the FDA recommends that measures should be *observational* rather than *proxy*. The guidance document states that these "observer reports ... include only those events or behaviors that can be observed." To illustrate this concept, the FDA guidance provides one example: "observers cannot validly report an infant's pain intensity but can report infant behavior thought to be caused by pain." Specific examples of observable signs of children's symptoms are provided in the FDA guidance [97] for the clinical development of drugs for the treatment of acute bacterial otitis media, an inflammation of the middle ear caused by a bacterial pathogen. For infants and young children who cannot articulate their symptoms, the guidance recommends assessing adult reports of observable signs such as ear tugging, ear rubbing, fussiness, and decreased appetite.

Proxy-report measures have yielded important information about child health and functioning, and this task force encourages future research involving parent proxy measures. Still, this task force agrees with the FDA recommendation that proxy measures should be avoided in research conducted for use in the regulatory context. Data contributing to regulatory decision making and medical product labeling must be as unambiguous as possible. Results based on proxy measures are associated with an unavoidable uncertainty, as indicated by frequent discrepancy between informant and child perspectives.

This task force also supports the FDA's recommendation for the use of observational measures whenever it is necessary to use informant reports in research conducted for medical product labeling. The discrepancy between parent and child reports has generally been shown to be greatest for nonobservable subjective domains, such as pain and emotional issues [9,25,98,99]. In contrast, these studies have frequently found stronger concordance between parent and child reports for more observable, objective domains such as physical functioning. These findings suggest that parent-report measures are likely to be most accurate and reliable when the items focus on observational content. Two examples of observational measures designed for the assessment of young children include the Keller Index of Nausea, which was designed for completion by pediatric nurses [100], and the parent-completed Revised Infant Gastroesophageal Reflux Questionnaire [101]. As illustrated in Figure 3, observational measures can focus on momentary assessment of the child's behavior completed at the time of observation or general

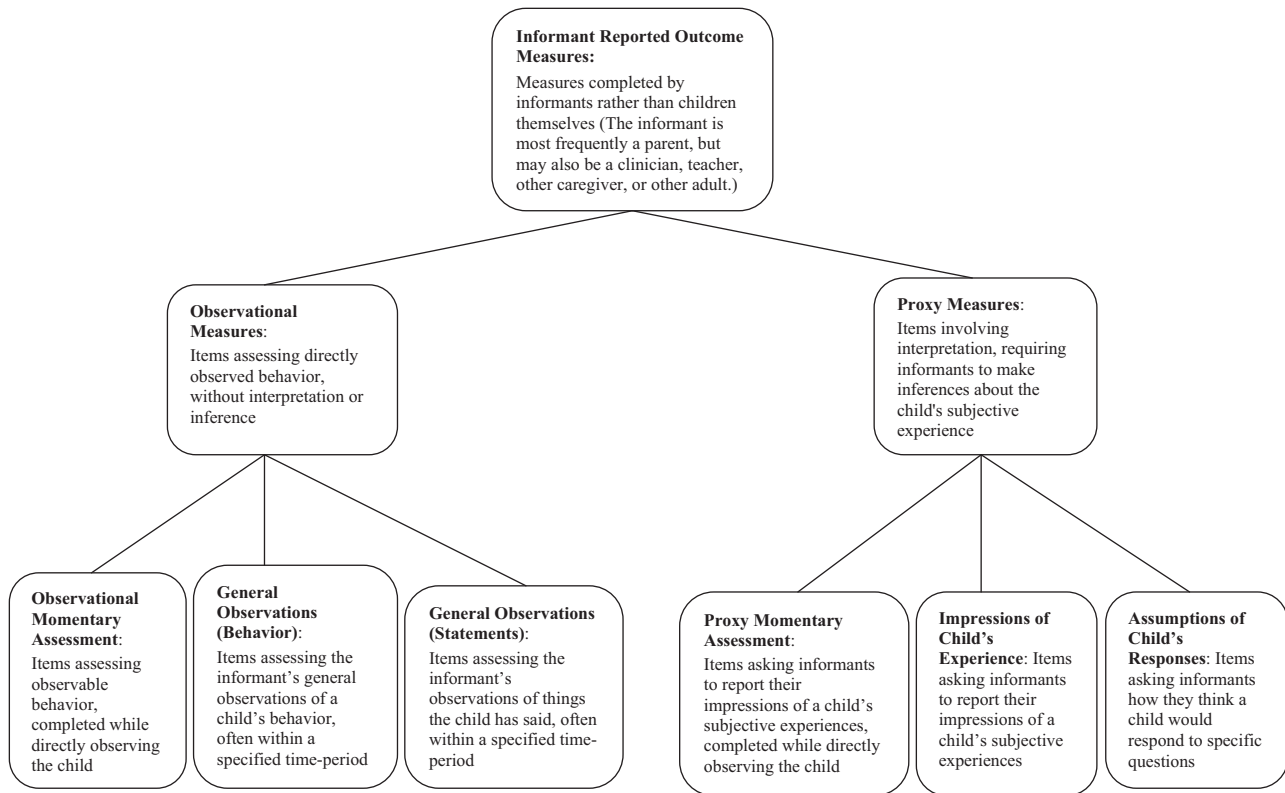


Fig. 3 – Types of informant-reported outcome measures for pediatric assessment.

observations of the child's behavior within a specified time period. Either of these two observational approaches may be appropriate for research intended to support medical product labeling. A third type of observational item, involving reports of statements the child has made, is usually less desirable because it depends on clarity of parent-child communication and accuracy of parents' memory regarding their children's statements.

Based on consideration of published literature and recent regulatory developments, this task force can provide three recommendations regarding the choice between child-reported and informant-reported outcome measures in the regulatory context. First, when children in the target age range are capable of completing a PRO instrument independently, a child-reported measure should be used (for discussion of developmental differences in ability to complete PRO measures, see the discussion of Good Research Practice 1). In the regulatory context, a child-reported measure is preferred because it is the most direct assessment of the child's experience of disease and treatment, without any bias or interpretation that could stem from informants' reports. To be considered capable of completing a PRO instrument, children within a target age range must demonstrate the ability to read a questionnaire, understand the relevant concepts, and provide reliable and valid responses. Reading and comprehension should be examined in cognitive interviews, while reliability and validity of responses can be documented in larger psychometric validation studies of PRO measures. Given the developmental differences within any age group, there will likely be some children who are not capable of completing a PRO measure even when the great majority of their peers are capable. Therefore, researchers may have to identify outliers and attend to general trends within a target age group when determining whether the age group should be considered capable of completing a particular PRO instrument.

Second, when children in the target age range are not capable of completing a PRO measure, an informant-reported measure may be used. Third, when using an informant-reported measure, all items in the instrument should assess observable content as much as possible, rather than subjective aspects of the child's experience.

It should also be noted that there are situations when it may be useful to have parents and children complete parallel forms of the same instrument. Interpretation of data from two simultaneous sources, however, often raises challenges as discussed in the "Conclusions and Recommendations for Future Research" section of the current report.

Good Research Practice 4: Ensure that the Instrument Is Designed and Formatted Appropriately for the Target Age Group

Several review articles have discussed aspects of pediatric PRO measure design and formatting that vary depending on the age of the respondents [6,8,9,14–16,24,102]. Partly in response to the emphasis on qualitative methods in the FDA PRO guidance, there is also a growing body of research directly assessing the child's perspective on questionnaire design and formatting [19,47,103,104]. These initial studies have yielded useful information, and our understanding of questionnaire design and formatting will continue to evolve as more pediatric cognitive interview studies are published. Therefore, we encourage more qualitative research focusing on ways to most effectively format and design PRO measures.

The following sections discuss a range of factors that must be considered to ensure that a PRO instrument is developmentally appropriate for the age range of a pediatric sample. It is important to consider these factors when developing a new instrument as

well as when implementing a previously validated instrument in a new age group. Specific age cutoffs for any aspect of instrument design depend on numerous interrelated factors such as item content and health-related characteristics of the target population. Therefore, this section does not provide age cutoffs that will apply in every situation. Instead, the purpose of this section is to review aspects of instrument design and format that are likely to differ among age groups, while providing recommendations based on developmental trends reported in published studies.

Health-related vocabulary and reading level

It has frequently been noted that children's health-related vocabulary increases with age [15,16]. Findings reported by Rebok et al. [19] illustrate this developmental trajectory. In this study, children's understanding of health-related terms was examined within four age groups: ages 5, 6, 7, and 8 to 11 years (Fig. 2). For most of the terms, the percentage of respondents with a poor understanding decreased as the age increased. For example, 14.3% of 5-year-old children have a poor understanding of the term "pain" compared with 8.3% of 6-year-olds, 6.2% of 7-year-olds, and 0% of older children. The percentages of poor understanding also decreased with age for other health-related terms, such as "energy," "healthy enough," "comfortable," and "nervous." Similar patterns of findings have been observed in other pediatric samples [71].

Clearly, it is critical that the vocabulary and reading level of a questionnaire are not more advanced than the linguistic abilities of the target population. Instrument developers have long been aware of this issue and have made efforts to simplify language and concepts in versions of questionnaires designed specifically for younger age groups [9,80,105]. One common approach is to assess the reading level of a questionnaire, using readability indices such as the Flesch-Kincaid Grade Level score, and edit the questionnaire until it is suitable for the target age range [9,80,106,107]. In addition to vocabulary, factors contributing to readability include syntax complexity, sentence length, and the number of sentences per paragraph [14]. One strategy for developing a questionnaire with age-appropriate vocabulary is to observe the language children use during concept elicitation interviews and focus groups [51]. Then, this vocabulary derived directly from children in the target age range can be incorporated into the resulting instrument.

When a pediatric PRO measure is used in a regulatory submission, documentation that the vocabulary level is fit for use in the target age group is important. Age appropriateness of reading level and vocabulary may be supported by readability index scores and data from cognitive interviews in which respondents from the target age group provide feedback on the comprehensibility of items. Cognitive interviewing methodology was found to be an effective approach for editing vocabulary of the PROMIS pediatric item bank [47]. In this qualitative interview study with 77 children and adolescents aged 8 to 17 years, cognitive interviews identified several terms that were incomprehensible to some children within this age group. Problematic terms included irritable (suggested change: cranky), exhausted (suggested change: tired), furious (suggested change: angry or mad), and social activities (suggested change: activities with friends). Similar results were reported in a study involving cognitive interviews focused on the PedsQL Diabetes Module [75]. Children reported that several words were difficult. Therefore, the version of the questionnaire for 8- to 12-year-old children was edited to include more age-appropriate language. For example, "fatigue" was changed to "tiredness," and "irritable" was changed to "cranky or grumpy."

Response scale

A review of health-related self-report measures for children aged 3 to 8 years found that Likert scales (i.e., scales with descriptors

along a continuum) [108,109] were the most commonly used type of response scale, appearing in 34 of the 53 measures identified for use in this age group [24]. Other types of scales included graphic representations, facial expressions, and a visual analogue scale, which is very rarely used in pediatric measures and frequently misunderstood by young children [110]. One study of 120 children aged 6 to 18 years found that children preferred Likert scales over visual analogue scales and numeric 10-point rating scales, reporting that Likert scales were easier to complete [104]. Furthermore, the cognitive interview study examining the PROMIS pediatric item bank found that even the youngest children in the sample (i.e., 8 years old) had no apparent difficulty understanding the response options of a five-point Likert scale [47]. In addition, a Delphi panel convened for the development of a child HRQOL measure yielded a consensus, with 94.1% of the respondents indicating that Likert scales were highly suitable for use with children [111].

In sum, Likert scales appear to be appropriate for most pediatric PRO measures, as indicated by qualitative research, expert opinion, children's opinions, and a substantial number of questionnaires that have used the scales in the past. However, Likert-scale items are not appropriate for all pediatric measures across age groups. Two studies have reported consistent findings regarding developmental differences in children's ability to use the full range of Likert scale response options. Younger children tend to provide more responses at the highest and lowest levels of the response scale. In one of the studies, 5- to 6-year-old children provided more extreme responses than did older children when rating emotion-based tasks on both three-point and five-point scales [44]. The other study found that the percentage of extreme responses decreased with age: 87.1% at age 5 years, 78.9% at age 6 years, 61.4% at age 7 years, and 50.4% at ages 8 to 11 years [19]. Findings from these two studies suggest that, for some items, Likert response scales may be appropriate for older, but not younger children. Therefore, we recommend evaluating the response options of a PRO measure with children of the target age group in cognitive interviews. In these interviews, it will be important to identify the age at which children begin to use and understand the full range of the response option scale, rather than relying on the extreme responses. To be considered adequate for PRO measures used in regulatory submissions, response options must also meet criteria described in the FDA PRO guidance [1], including clear distinction between response choices, clear wording, and adequate instructions (possibly including an example of a completed item).

Recall period

A recent review of instruments assessing children's HRQOL reported that many commonly used child-reported measures have a recall period of the past 4 weeks or the past month [14]. These instruments include the CHIP [79,112], the Child Health Questionnaire [15], and the PedsQL [113]. Measures, such as these three, with relatively long recall periods have clearly yielded important information about children's health in many clinical trials and other studies. However, instruments with recall periods as long as 1 month may be met with skepticism if used as the basis for a claim in medical product labeling.

The FDA PRO guidance [1] says,

PRO instruments that call for patients to rely on memory, especially if they must recall over a long period of time ... are likely to undermine content validity. Response is likely to be influenced by the patient's state at the time of recall. For these reasons, items with short recall periods or items that ask patients to describe their current or recent state are usually preferable.

The FDA PRO guidance does not provide an optimal recall period because the choice of the recall period depends on a wide range of factors, including the instrument's purpose as well as the duration, frequency, and intensity of the concept being measured. The guidance adds that the appropriateness of a recall period should be evaluated with regard to patients' "ability to validly recall the information requested."

There is some indication that children's ability to accurately use recall periods varies among child age groups. For example, one study found that children younger than 8 years were less likely than older children to be able to report the last time a symptom occurred in the past 4 weeks, and it was not clear that 5- and 6-year-old children understood the concept of a week or a month [19]. A study with parent-child dyads also highlighted children's limitations in using recall periods [71]. In this study, children sometimes appeared to lack the cognitive skills necessary for comprehending time frames, and parents often helped their children remember health-related difficulties by linking the time frame to specific events and activities. Results were not presented by age, but 69% of the sample was between 8 and 10 years old, suggesting that some children within this age range may have difficulty with recall periods. Another recent study examined the effect of "retention interval" (i.e., elapsed time between the assessment and the event to be reported) on the accuracy of children's memory for meals [114]. In this sample of fourth-grade children (i.e., roughly 9 years old), the accuracy of memory of school breakfasts and lunches was improved by shortening the retention interval. This finding highlights the importance of using shorter recall periods for child-reported measures. PRO instrument developers are likely aware that children could have more difficulty with recall periods than do adults. For example, the Child Asthma Quality of Life Questionnaire has a recall period of 1 week, which was reduced from 2 weeks in the adult version of the instrument [115].

In sum, shorter recall periods are preferable for PRO measures used in the regulatory context, and this may be more important for pediatric measures than for adult measures. This task force believes that recall periods of 24 hours or less are likely to be viewed more favorably than longer recall periods in the regulatory context based on the FDA PRO guidance recommendation for shorter recall periods and research demonstrating children's difficulties comprehending longer time frames. Furthermore, a momentary assessment approach (i.e., assessing the child's perceptions of his or her current state) may be useful because it avoids any problems that could be caused by children's understanding of a period of time.

It is not possible, however, to provide a specific recall period that will be optimal for all pediatric studies, and longer recall periods may be acceptable in some situations if adequate justification is provided. Shorter recall periods may also have disadvantages, such as the need for more frequent measurement and the possibility that assessments could fail to capture important health-related symptoms or events that occur outside the specified recall period. In addition, it may be advantageous for data collection and interpretation for PRO measures to have recall periods that are consistent with clinical measures in a particular trial. Some therapeutic areas such as psychiatry tend to use clinician-reported measures with recall periods of 1 week, and a PRO with a shorter recall period would introduce inconsistency that could interfere with interpretation of findings. All these factors will need to be considered when determining the appropriate recall period, and cognitive interviews are likely to be important in justifying a recall period for each purpose, instrument, and age group. To inform future decisions about recall periods, research is needed on the accuracy of children's memory for health-related concepts across various time frames, and these studies should be conducted within narrow age groupings.

Length of instrument

The length of a pediatric PRO questionnaire requires careful consideration because of the wide variation in children's ability to maintain attention to tasks [16]. It is generally believed that instruments designed for use in younger children face more length limitations than do those designed for use in older children and adults. It has also been suggested that younger children require shorter instructions than do older children [47]. Measures that are overly long may cause children to omit items or think less carefully about each item, thus yielding less accurate and reliable data [9]. In consideration of these issues, one approach has been to create multiple versions of a single instrument, varying in length and targeted at different age groups [105]. Pilot testing and cognitive interviews in children of the target age range will help determine whether the length of the instrument is appropriate for the target age range, prior to including the instrument in a clinical trial.

Pictorial representations

Instrument developers have often included pictorial representations of concepts and response options in pediatric PRO measures [14]. There is little empirical data demonstrating the advantages of these illustrations, although one study did find that a pediatric questionnaire with illustrations was significantly faster to complete than a version without illustrations, suggesting that the cartoon version may engage the child more effectively [116]. Despite minimal empirical support, many researchers believe that these pictorial approaches may help younger children complete questionnaires. It is theorized that pictures can help maintain children's interest, sustain attention, increase item comprehensibility, clarify the response process, and therefore, foster more meaningful responses [9]. More research is needed to examine and confirm these potential benefits, but based on the opinions of experts and the many instruments that have used this approach, it seems likely that pictures may be helpful for some younger children.

Several pictorial approaches have been used. The simplest approach to make response options more concrete may be to use circles of increasing sizes. For example, the CHIP-CE requires children to respond to items by choosing one of several response options [112]. There is a circle above each response, and these circles increase in size as the response options progress from "never" to "always" (Fig. 4). Response options may also be represented by simple drawings of faces, with expressions that indicate neutrality or various degrees of positive or negative emotion (Fig. 5) [105,117–119]. This "smiley face" approach has been recommended for use with young children by 80% of the participants in a Delphi panel conducted during the development of the KIDSCREEN quality-of-life measure [111].

There are also examples of questionnaires using more elaborate drawings to illustrate items and response options. The CHIP-CE has illustrations of children representing the extreme responses of the Likert response scale for each item (Fig. 4) [112]. Other measures have used cartoon-like illustrations intended to help children understand item content, including drawings of children representing psychiatric diagnoses (Fig. 6) [120] and cartoons of dogs representing pediatric dermatological symptoms [116].

Results of one study, however, suggest that researchers should be cautious when using these pictorial approaches. Chambers and Craig [118] compared two types of pain rating scales using faces as response options in a sample of 100 children. One of the scales included a neutral face as the "no pain" anchor, whereas the other scale had a smiling face as the "no pain" anchor (Fig. 7). The two types of scales yielded significantly different item scores, with the direction of difference

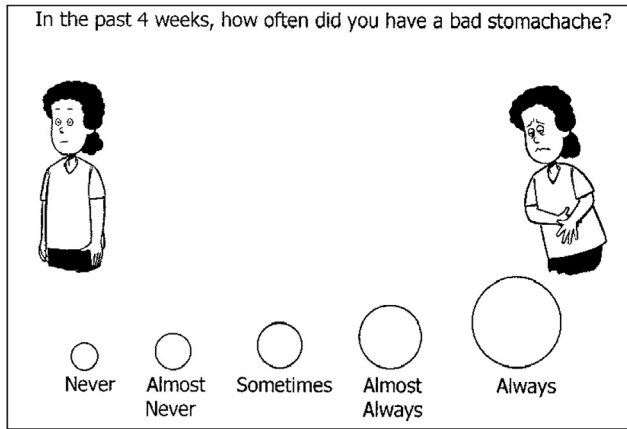


Fig. 4 – Example of a response scale using illustrations and circles of increasing sizes to clarify response options. This item is presented here as an example of a PRO item using pictorial representations to help children understand the response scale. However, other aspects of this item are not optimal for use in the regulatory context. For example, the recall period is relatively long, and the response options require the child to average their experience of the symptom over this 4-week period. This is a complex averaging task, and children’s memory is unlikely to be accurate for a 4-week period. Therefore, regulatory reviewers may have an unfavorable view of this item. PRO, patient-reported outcome. Reprinted from *Med Care*, 42, Riley AW, Forrest CB, Rebok GW, et al., *The Child Report Form of the CHIP-Child Edition: reliability and validity*, 221–31, 2004, with permission from Lippincott Williams & Wilkins.

depending on the content of the item. This finding indicates that illustrations must be designed and tested as rigorously as other aspects of the questionnaire because these illustrations could confound results and undermine validity. Another possible risk of illustrations is that some adolescents could perceive them as overly childish. As with other aspects of questionnaire design, cognitive interviews can be used to examine children’s perceptions and understanding of pictorial representations.

Other formatting details

Other formatting details that may initially seem less important could also have an impact on the clarity and validity of a pediatric PRO measure. Maintaining a clear layout of items with sufficiently large print may enhance readability, particularly for younger children. These types of formatting issues are rarely discussed in the published literature, but there is some indication

from previous research that these are not trivial concerns. For example, the format of the Child Oral Health Impact Profile was revised with increased font size and shading for every other item as a result of qualitative interviews [103]. Furthermore, in the cognitive interviewing study for the PROMIS pediatric item bank, child respondents suggested making the font larger and stating the recall period in bold type for additional clarity [47].

Administration approaches

When developing and implementing adult PRO measures, researchers can generally assume that most patients can complete questionnaires independently. However, children differ in their ability to independently complete questionnaires. Therefore, when developing pediatric PRO measures, researchers will need to consider a variety of administration approaches that differ in the degree of independence expected of a child respondent. Older children and adolescents can often be expected to complete written questionnaires independently. Younger children may require interviewer-administered measures or parental assistance to overcome limitations in reading ability and attention. For example, the Childhood Asthma Questionnaire was developed with three forms, corresponding to three age ranges [105,119]. The form for the youngest age range (4–7 years old) is completed by the child with a parent’s help, the form for the middle age range (8–11 years old) may be completed with adult help if necessary, and the form for the older age group (12–16 years old) was designed to be completed independently (see “Conclusions and Recommendations for Future Research” later in this report for discussion of challenges combining data from multiple age-specific versions of an instrument). Other types of administration approaches, such as electronic data collection (see below) or the use of props, may also help enhance children’s attention and comprehension during the measure completion process [24].

The choice of administration approach will depend on several factors, including the children’s age and the complexity of the constructs being assessed. The administration approach should be considered carefully during the qualitative phase of instrument development. For example, if cognitive interviews suggest that children are having difficulty completing a measure independently, perhaps a less independent administration approach would yield more valid data. As discussed in *Good Research Practice 2*, however, the presence of a parent or another adult could bias or inhibit the child’s responses. Therefore, this task force recommends that parents and other adults should assist children with PRO instrument completion only when truly necessary.

To ensure successful data collection, it is essential that project staff members who administer interviews and other study measures are adequately trained. Training is particularly important for open-ended procedures such as concept elicitation interviews,

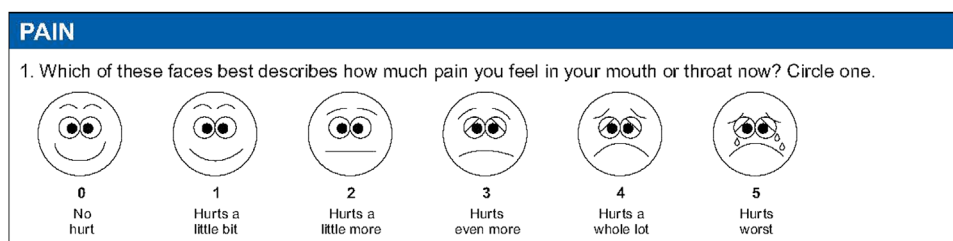


Fig. 5 – Example of a response scale with facial expressions illustrating each response option.

Reprinted from *Eur J Oncol Nurs*, 14, Tomlinson D, Gibson F, Treister N, et al., *Refinement of the Children’s International Mucositis Evaluation Scale (ChIMES): child and parent perspectives on understandability, content validity and acceptability*, 29–41, 2010, with permission from Elsevier.

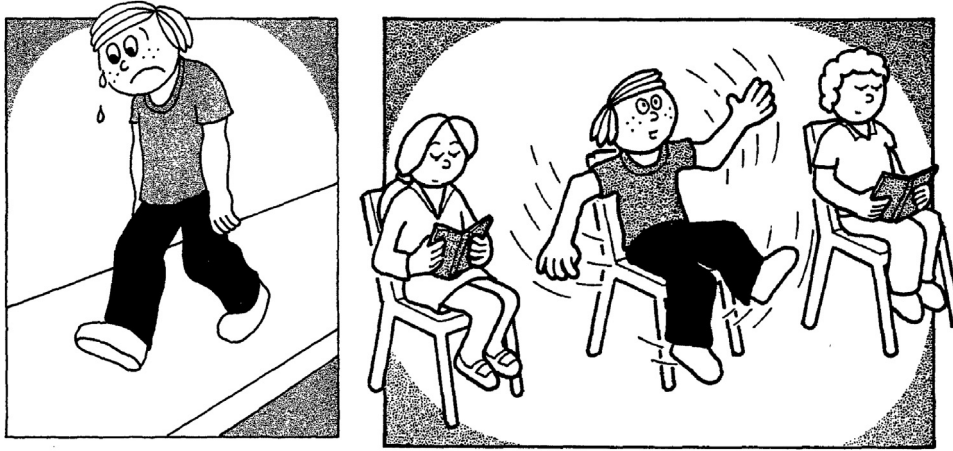


Fig. 6 – Examples of figures that have been used to illustrate psychiatric disorders in a child-reported questionnaire. Reprinted from J. Abnorm Child Psychol, 22, Valla JP, Bergeron L, Berube H, et al., A structured pictorial questionnaire to assess DSM-III-R-based diagnoses in children (6-11 years): development, validity, and reliability, 403-23, 1994. With kind permission from Springer Science and Business Media.

which require greater sensitivity and insight from the interviewer. The process of child and adolescent interviews may differ from that of adult interviews. Interviewers must have some basic knowledge of developmental issues as well as strategies for communicating with children and reading behavioral cues [121]. Recommendations for interviewing parents and teachers about children's behavior and background have also been published [121].

Although this task force report focuses specifically on PRO administration, it should be noted that research with children requires careful attention to all aspects of the data collection process. For example, researchers will need to develop age-appropriate language for explaining the study purpose and procedures to children. Before children begin completing PRO measures, interviewers or other study staff members should inform them of the following: what will be required of them, the purpose of the questions that will be asked, the intended use of the data, confidentiality procedures, and what to do if they become uncomfortable or want to stop participating. While some children will not comprehend the study details as well as do adults, researchers should make an effort to give children a general understanding and ensure that they feel comfortable.

Electronic data collection

Electronic approaches to collecting PRO data (i.e., ePRO) may have significant benefits in research with children and adolescents. Screen-based platforms that may be useful with children include handheld devices such as smart phones, tablets and touch screens that operate without the use of a keyboard or mouse, and desktop or laptop computers [122]. Several published studies suggest that electronic questionnaire assessment, including Internet-based administration, with children can be a feasible, reliable, and valid data collection approach [123-127]. Many children are familiar and comfortable with screen-based activities, and one study found that children preferred Internet administration of a health and behavior questionnaire over a paper version [123]. If screen-based modes of administration can help children stay focused and engaged, they may help to improve the quality of self-reported data while minimizing missing data. Interactive voice response technology may also be a feasible data collection approach for some age groups [128]. This convenient approach involving verbal presentation of questions via telephone may be easier to process than written questionnaires for some children with reading or writing limitations.

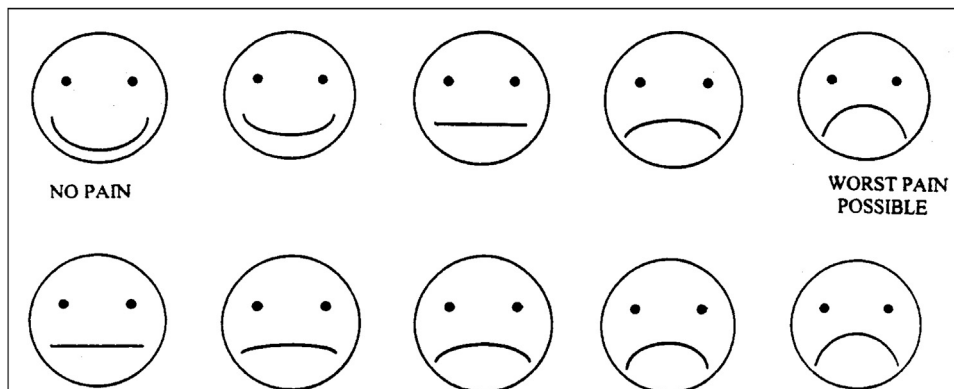


Fig. 7 – Two types of face scales with different anchors for “no pain.” Reprinted from Pain, 78, Chambers CT, Craig KD, An intrusive impact of anchors in children's faces pain scales, 27-37, 1998, with permission from the International Association for the Study of Pain.

Previous ISPOR PRO task forces have provided recommendations for development and testing of ePRO measures, including cognitive interviews, usability testing, and evaluation of equivalence between paper and electronic versions of questionnaires [52,122]. These previous recommendations for ePRO development are relevant to measures for children and adolescents. In addition, the current task force recommends usability assessment of electronic data collection methods within narrow age groupings when developing ePRO measures for children. Like self-reported written questionnaire administration, screen-based technology relies on the respondent's reading ability, and ePRO administration may be effective in one age group, but not in others. Therefore, researchers will need to document usability across all applicable age groups for ePRO measures to be considered useful in the regulatory context.

Good Research Practice 5: Consider Cross-Cultural Issues

When developing pediatric instruments for widespread use, it is important to consider potential differences associated with culture and language. Although methods for the translation and cultural validation of pediatric measures are similar to those used for adult measures [129–131], pediatric measures may raise additional issues that need to be considered. For example, because of differences in educational systems across countries, the reading ability of children at any given age may vary. Differences in script among diverse languages may also lead to geographical variation in literacy within specific age groups [132]. Therefore, it may be necessary to reassess an instrument's age-appropriateness in each new country where it may be administered to children. There are also likely to be cultural differences in the type of information that is conveyed to children about disease and treatment, as well as differences in the degree to which children are regarded as independent reporters of their well-being. In addition, there may be cultural differences in children's willingness to talk to interviewers without a parent present. Such cultural norms will need to be examined as part of establishing content validity within each new country. Finally, the impact of disease and treatment on social, emotional, and role functioning may vary greatly across cultures, depending on the typical activities of children in the culture.

Therefore, the appropriateness of a pediatric PRO instrument will need to be reexamined with content validity assessment within each new culture. This assessment should focus on all relevant aspects of the instrument including the instructions, items, concepts, vocabulary, and pictorial representations. In sum, cross-cultural PRO instrument development for children is likely to require greater sensitivity and effort than simply following the cross-cultural guidelines set forth for adult instruments. Future research and guidelines may help to clarify specific steps that must be taken to translate and validate pediatric PRO instruments across cultures.

Conclusions and Recommendations for Future Research

The recommendations provided in this task force report should not be interpreted as concrete rules that must be followed in every study. Instead, the intention of this task force was to present general guidance and discuss the important issues that must be considered when designing, validating, or implementing pediatric PRO instruments for use in the context of regulatory submissions and medical product labeling. Each individual study is likely to be unique, and the optimal measurement approach will depend on a range of factors, including the child's age, the medical condition of the target population, and the constructs being assessed.

Furthermore, different PRO approaches and study designs may be advisable for research that is not intended to support medical product labeling or regulatory decisions. Therefore, the good research practices recommended in the current report should be viewed as a starting point, and each researcher will have to consider the purpose and context of each individual study.

PRO research with children and adolescents is a relatively young area of research, and more work is needed to provide updated PRO instruments and methodological guidance for future studies. For example, research on optimizing PRO design for younger children is needed, particularly for children younger than 8 years for whom self-reported measures often have inconsistent reliability and validity. There may be specific ways to phrase, present, or format items that will maximize comprehensibility for young children, thus enhancing instrument reliability and validity. Studies comparing multiple measurement approaches may help provide more specific recommendations than are currently available. In addition, more work is needed to identify the qualitative research strategies that are most likely to be effective with younger children. The current report provides initial recommendations for qualitative methods, but there are not many published studies that have examined the content validity of PRO instruments for children. Therefore, additional research is needed to examine and refine these methods. As more qualitative research with this population is conducted and published, the methodology is likely to evolve.

When informant-reported outcome instruments must be used, there is a growing emphasis on developing truly observational items, rather than proxy measures that require inference into the child's subjective experience. Therefore, it may be useful to update and validate commonly used parent-reported and clinician-reported instruments to reflect this more observational approach.

Another challenge involves the interpretation of data from multiple age groups. Many PRO measures for children are developed with multiple versions for different age groups. Item content, response scales, vocabulary, pictorial representations of response options, and other instrument characteristics may vary across the age-specific versions to ensure that each version is appropriate for a specific age range. The reporter may also vary, with parents reporting for younger children while older children complete questionnaires themselves. This "multiple versions" approach can ensure that a particular instrument examines age-appropriate content with age-appropriate assessment procedures throughout a broad age range. When PRO research is conducted in the regulatory context, however, the goal is often to support a single labeling claim, rather than individual claims for each narrow age group. Therefore, the use of multiple versions raises the question of how data from different versions or even different reporters can support a single claim across childhood. One possible analytic approach is to pool data from multiple versions in a single analysis, but it would be necessary to demonstrate that content and measurement properties of all versions are truly equivalent and comparable. Another approach is to analyze different versions separately and examine whether treatment outcome trends are similar across age ranges and versions of the instrument. Future research is needed to provide guidance on strategies for analyzing and interpreting data gathered with multiple age-specific versions of a single PRO instrument.

It can also be challenging to interpret data from parents and children completing parallel forms of the same instrument. In some situations, both parents and children may be able to provide useful information, but it is not clear how results should be interpreted if the two reporters have different perceptions of the same construct. Furthermore, with increasing emphasis on limiting informant-reported measures to observable domains (as discussed in Good Research Practice 3), some parent-reported

forms could have slightly different content than do child-report forms, which may include more subjective content. Therefore, when gathering data from both parents and children, it may be best to specify *a priori* that data from one of these reporters will be considered the primary end point. More research is needed to provide guidance for taking both children's and parents' perspectives into account, even when they diverge.

Additional research is also needed to examine the responsiveness of pediatric PRO instruments. Responsiveness is the extent to which a health status measure accurately detects change in a patient's condition over time [133–135]. In the regulatory context, responsiveness is necessary for a PRO measure to be considered fit for the purpose of "identifying differences in scores over time in both individuals and groups who have changes with respect to the measured concept" [1]. As reported in the discussion of Good Research Practice 1 in the current report, reliability and validity have been analyzed and reported for many PRO measures designed for children and adolescents. Relatively little data on responsiveness of pediatric PRO measures, however, have been published [14]. Given the importance of accurately detecting change in studies conducted to support medical product labeling, this task force recommends evaluating the responsiveness of any pediatric PRO measure used in this context.

Although this task force report focuses primarily on the development of new PRO instruments, the recommendations and discussion can also apply to the use, evaluation, and modification of existing PRO measures designed for use in children. Issues pertaining specifically to the use and modification of existing instruments have been discussed in detail by a previous ISPOR task force [54].

One limitation of this task force was the narrow focus on pediatric PRO assessment in the context of research conducted for regulatory submissions. Health outcomes research outside this regulatory context was considered to be outside the scope of this report, but it should be noted that other areas of research on children and adolescents also raise important methodological challenges. For example, health state utility valuation is necessary for deriving quality-adjusted life years (QALYs) in cost-utility modeling submitted to reimbursement authorities and other decision makers. However, valuing childhood health states raises unique challenges [136–138]. Children may not be able to conceptualize the abstract concepts necessary for direct utility assessment such as time trade-off and standard gamble tasks, and common utility assessment methods used for adults often do not apply to pediatric health states. Furthermore, given the rapid physical and cognitive development of childhood, child health states are less likely to remain stable than adult health states, and evaluation of temporary health states is complex [139]. Still, cost-utility analyses are frequently conducted to model treatment of childhood health conditions, and utilities for pediatric health states are necessary to quantify children's quality of life for these models [140]. Consequently, researchers have been working to develop health utility measures specifically for children and adolescents [141–144]. Another growing area of research with the potential to influence clinical and policy decision making is comparative effectiveness research (CER). It has been suggested that methods and concepts of CER may differ when examining treatments for children rather than adults [145]. Because of the specific focus on PROs in the regulatory context, the current task force did not address these areas of research, but it is hoped that future research will continue to examine child-specific issues in a broad range of contexts such as health state utility valuation and CER.

Several other important topics related to pediatric research were not covered in the current report. For example, evaluation of drugs in children raises unique ethical issues, which are discussed in detail in the published guidelines from the American Academy of Pediatrics [146]. Issues related to the recruitment of children

for randomized clinical trials have also been discussed previously [147].

With additional PRO studies in children and adolescents, confidence in pediatric PRO instruments and methodology will continue to grow. Furthermore, as PRO data are increasingly used to support labeling claims, there will be more information regarding the standards by which these instruments will be judged. As this area of research continues to develop, the use of PRO instruments in clinical trials and regulatory submissions will help ensure that children's experience of disease and treatment is accurately represented and considered in regulatory decisions.

Acknowledgments

The authors thank Elizabeth Molsen for her endless energy, enthusiasm, and patience with our debates, delays, and strong personalities. This report would not exist without Elizabeth's support.

The authors thank the following reviewers for valuable feedback on earlier drafts of this report: Bhagwan Aggarwal, Celestina Arrigo, Alexandra Barsdorf, Katy Benjamin, Andrew Bottomley, Yan Cheng, Hsingwen Chung, Ben Craig, Dalia Dawoud, Prasanna Deshpande, Pero Draganić, Ancilla Fernandes, Emmuela Flood, Mary Gawlicki, Richa Goyal, Rosina Hinojosa, Stefan Holmstrom, Manthan Janodia, Peter Kaskel, David Keleti, Kanav Khera, Ramanath Kv, Jeanne Landgraf, Kathryn Lasch, Tamar Lasky, Yvonne Lis, Michelle Mocarski, Chris Morris, Shekoufeh Nikfar, Mira Pavlovic, Farhan Abdul Rauf, Yong Joo Rhee, Margaret Rothman, Joerg Ruof, Steve Ryder, Luciana Scalone, Jun Su, Tara Symonds, Elio Tanaka, Sreedhar Tirunagari, Nayanabhirama Udupa, Wendy Ungar, Erica Velthuis, and Etta Vinik.

The authors also thank four members of regulatory and reimbursement agencies who provided very useful comments on earlier drafts of this report: Laurie B. Burke, RPh, MPH, Director, Study Endpoints and Labeling Development, Office of New Drugs, CDER, FDA, USA; Andrew E. Mulberg, MD, FAAP, Deputy Division Director, Gastroenterology and Inborn Errors Products, Office of New Drugs, CDER, FDA, USA; Melissa S. Tassinari, PhD, DABT, Senior Clinical Analyst, Pediatric and Maternal Health Staff, Office of New Drugs, CDER, FDA, USA; Andreas Gerber-Grote, PhD, MD, Head of Health Economics, Institute for Quality and Efficiency in Health Care (IQWiG), Germany; and Mira Pavlovic, DrPH, Deputy Director for Health Technology Assessment, Haute Autorité de Santé, France, and Joint Action Member, EUnetHTA.

The authors thank Sonya Eremenco for consultation on the ePRO section and Margaret Vernon for consultation on the content validity section.

The first author thanks several additional people who made this report possible: Thanks to Nancy Kline Leidy for her faith in me and for connecting me to this project. Thanks to Kate Van Brunt and Shannon Kummer for literature searching. Thanks to Laurie Smith, Candice Norton, Natalie Washington, Tamaqua Nisbett, and Dana Christiansen for tenaciously tracking down all the references cited in this report, as well as many other references that were not cited. Thanks to Kelly McDaniel for truly thoughtful editing and proofreading. Thanks to Fritz Hamme for copy editing. Thanks to Jessica Jordan and Aria Gray for their never-ending loyalty and help with too many aspects of this report to mention here. Thanks to Week 52 for giving me the time to write. Most importantly, thanks to Sonya and Freya for their understanding on the nights when this task force report kept me occupied or preoccupied (Freya was 8 months old when this task force began, and she is now almost old enough to complete a PRO measure, although it would have to be interviewer-administered and she might require a ripe mango to stay focused and motivated).

Source of financial support: None of the authors received financial support for their participation in this task force. All authors volunteered their time for discussion, research, and writing of this report.

REFERENCES

- [1] Food and Drug Administration. Guidance for Industry—Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Silver Spring, MD: Food and Drug Administration, 2009.
- [2] Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health* 2007;10(Suppl. 2):S125–37.
- [3] Burke LB, Kennedy DL, Miskala PH, et al. The use of patient-reported outcome measures in the evaluation of medical products for regulatory approval. *Clin Pharmacol Ther* 2008;84:281–3.
- [4] Gnanasakthy A, Mordin M, Clark M, et al. A review of patient-reported outcome labels in the United States: 2006 to 2010. *Value Health* 2012;15:437–42.
- [5] European Medicines Agency. Reflection Paper on the Regulatory Guidance for the Use of Health-Related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products. London, United Kingdom: European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP), 2005.
- [6] Connolly MA, Johnson JA. Measuring quality of life in paediatric patients. *Pharmacoeconomics* 1999;16:605–25.
- [7] Droter D. Measuring Health-Related Quality of Life in Children and Adolescents: Implication for Research and Practice. Mahwah, NJ: Erlbaum, 1998.
- [8] Eiser C, Mohay H, Morse R. The measurement of quality of life in young children. *Child Care Health Dev* 2000;26:401–14.
- [9] Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. *Health Technol Assess* 2001;5:1–157.
- [10] Eiser C, Morse R. A review of measures of quality of life for children with chronic illness. *Arch Dis Child* 2001;84:205–11.
- [11] Erling A. Methodological considerations in the assessment of health-related quality of life in children. *Acta Paediatr Suppl* 1999;88:106–7.
- [12] De Civita M, Regier D, Alamgir AH, et al. Evaluating health-related quality-of-life studies in paediatric populations: some conceptual, methodological and developmental considerations and recent applications. *Pharmacoeconomics* 2005;23:659–85.
- [13] Hinds PS, Brandon J, Allen C, et al. Patient-reported outcomes in end-of-life research in pediatric oncology. *J Pediatr Psychol* 2007;32:1079–88.
- [14] Landgraf JM. Practical considerations in the measurement of HRQoL in child/adolescent clinical trials. In: Fayers P, Hays R, eds., *Assessing Quality of Life in Clinical Trials*. (2nd ed.). New York: Oxford University Press, 2005.
- [15] Landgraf JM, Abetz LN. Measuring health outcomes in pediatric populations: issues in psychometrics and application. *Qual Life Pharmacoecon Clin Trials* 1996;2:793–802.
- [16] Matza LS, Swensen AR, Flood EM, et al. Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. *Value Health* 2004;7:79–92.
- [17] Morris C, Gibbons E, Fitzpatrick R. Child and Parent Reported Outcome Measures: A Scoping Report Focusing on Feasibility for Routine Use in the NHS. Oxford, United Kingdom: Department of Public Health, University of Oxford, 2009.
- [18] Quittner AL, Modi A, Cruz I. Systematic review of health-related quality of life measures for children with respiratory conditions. *Paediatr Respir Rev* 2008;9:220–32.
- [19] Rebok G, Riley A, Forrest C, et al. Elementary school-aged children's reports of their health: a cognitive interviewing study. *Qual Life Res* 2001;10:59–70.
- [20] Theunissen NC, Vogels TG, Koopman HM, et al. The proxy problem: child report versus parent report in health-related quality of life research. *Qual Life Res* 1998;7:387–97.
- [21] Varni 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 4.0 Generic Core Scales. *Health Qual Life Outcomes* 2007;5:1.
- [22] Wallander JL, Schmitt M, Koot HM. Quality of life measurement in children and adolescents: issues, instruments, and applications. *J Clin Psychol* 2001;57:571–85.
- [23] Clarke SA, Eiser C. The measurement of health-related quality of life (QOL) in paediatric clinical trials: a systematic review. *Health Qual Life Outcomes* 2004;2:66.
- [24] Cremeens J, Eiser C, Blades M. Characteristics of health-related self-report measures for children aged three to eight years: a review of the literature. *Qual Life Res* 2006;15:739–54.
- [25] Eiser C, Morse R. Can parents rate their child's health-related quality of life? Results of a systematic review. *Qual Life Res* 2001;10:347–57.
- [26] Rajmil L, Herdman M, Fernandez de Sanmamed MJ, et al. Generic health-related quality of life instruments in children and adolescents: a qualitative analysis of content. *J Adolesc Health* 2004;34:37–45.
- [27] Ravens-Sieberer U, Erhart M, Wille N, et al. Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. *Pharmacoeconomics* 2006;24:1199–220.
- [28] Solans M, Pane S, Estrada MD, et al. Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments. *Value Health* 2008;11:742–64.
- [29] Bevans KB, Riley AW, Moon J, et al. Conceptual and methodological advances in child-reported outcomes measurement. *Expert Rev Pharmacoecon Outcomes Res* 2010;10:385–96.
- [30] Riley AW. Evidence that school-age children can self-report on their health. *Ambul Pediatr* 2004;4:371–6.
- [31] Williams C. Gender, adolescence and the management of diabetes. *J Adv Nurs* 1999;30:1160–6.
- [32] Patrick DL, Edwards TC, Skalicky AM, et al. Validation of a quality-of-life measure for deaf or hard of hearing youth. *Otolaryngol Head Neck Surg* 2011;145:137–45.
- [33] Peden M. World report on child injury prevention appeals to “Keep Kids Safe”. *Inj Prev* 2008;14:413–4.
- [34] Office of the United Nations High Commissioner for Human Rights. Convention on the Rights of the Child. Geneva, Switzerland: United Nations, 1989.
- [35] Centers for Disease Control and Prevention. Protect the Ones You Love: Child Injuries Are Preventable. Atlanta, GA: Centers for Disease Control and Prevention, 2012.
- [36] Borse NN, Rudd RA, Dellinger AM, et al. Years of Potential Life Lost Due to Unintentional Injuries among Children and Adolescents. Atlanta, GA: Centers for Disease Control and Prevention, 2012.
- [37] US Department of Health & Human Services. What Is Adolescence? Washington, D.C.: US Department of Health & Human Services, Office of Population Affairs, 2012.
- [38] Kaplan PS. Adolescence. Boston: Houghton Mifflin Company, 2004.
- [39] Grange A, Bekker H, Noyes J, et al. Adequacy of health-related quality of life measures in children under 5 years old: systematic review. *J Adv Nurs* 2007;59:197–220.
- [40] Khan FI, Reddy RC, Baptist AP. Pediatric Dyspnea Scale for use in hospitalized patients with asthma. *J Allergy Clin Immunol* 2009;123:660–4.
- [41] Luby JL, Belden A, Sullivan J, et al. Preschoolers' contribution to their diagnosis of depression and anxiety: uses and limitations of young child self-report of symptoms. *Child Psychiatry Hum Dev* 2007;38:321–38.
- [42] Riley A, Forrest CB, Starfield B, et al. Technical Manual for the Child Health and Illness Profile-Child Edition and CHIP-CE. Baltimore, MD: The Johns Hopkins University, 2001.
- [43] Ronen GM, Streiner DL, Rosenbaum P, et al. Health-related quality of life in children with epilepsy: development and validation of self-report and parent proxy measures. *Epilepsia* 2003;44:598–612.
- [44] Chambers CT, Johnston C. Developmental differences in children's use of rating scales. *J Pediatr Psychol* 2002;27:27–36.
- [45] Demetriou A, Raftopoulos A. Modeling the developing mind: from structure to change. *Dev Rev* 1999;19:319–68.
- [46] Juniper EF, Guyatt GH, Feeny DH, et al. Minimum skills required by children to complete health-related quality of life instruments for asthma: comparison of measurement properties. *Eur Respir J* 1997;10:2285–94.
- [47] Irwin DE, Varni JW, Yeatts K, et al. Cognitive interviewing methodology in the development of a pediatric item bank: a Patient Reported Outcomes Measurement Information System (PROMIS) study. *Health Qual Life Outcomes* 2009;7:3.
- [48] Edwards TC, Huebner CE, Connell FA, et al. Adolescent quality of life, part I: conceptual and measurement model. *J Adolesc Health* 2002;25:275–86.
- [49] Arnett JJ. Emerging adulthood: a theory of development from the late teens through the twenties. *Am Psychol* 2000;55:469–80.
- [50] Henig RM. What is it about 20-somethings? *The New York Times*. August 22, 2010:MM28.
- [51] Patrick DL, Burke LB, Gwaltney CJ, et al. 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* 2011;14:967–77.
- [52] Patrick DL, Burke LB, Gwaltney CJ, et al. 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* 2011;14:978–88.

- [53] Fehnel S. Establishing optimal requirements for content validity: a work in progress. *Value Health* 2009;12:1074.
- [54] Rothman M, Burke L, Erickson P, et al. 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* 2009;12:1075–83.
- [55] Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res* 2009;18:1263–78.
- [56] Leidy NK, Vernon M. Perspectives on patient-reported outcomes: content validity and qualitative research in a changing clinical trial environment. *Pharmacoeconomics* 2008;26:363–70.
- [57] Collins D. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res* 2003;12:229–38.
- [58] Deane DM. Content, construct, and criterion-related validity. *Diabetes Educ* 1991;17:361–2.
- [59] Gannotti ME, Cruz C. Content and construct validity of a Spanish translation of the Pediatric Evaluation of Disability Inventory for children living in Puerto Rico. *Phys Occup Ther Pediatr* 2001;20:7–24.
- [60] McCabe MA, Granger CV. Content validity of a pediatric functional independence measure. *Appl Nurs Res* 1990;3:120–2.
- [61] Penny AM, Waschbusch DA, Klein RM, et al. Developing a measure of sluggish cognitive tempo for children: content validity, factor structure, and reliability. *Psychol Assess* 2009;21:380–9.
- [62] Morris C, Liabo K, Wright P, et al. Development of the Oxford Ankle Foot Questionnaire: finding out how children are affected by foot and ankle problems. *Child Care Health Dev* 2007;33:559–68.
- [63] Riesch SK, Anderson LS, Angresano N, et al. Evaluating content validity and test-retest reliability of the children's health risk behavior scale. *Public Health Nurs* 2006;23:366–72.
- [64] Stewart JL, Lynn MR, Mishel MH. Evaluating content validity for children's self-report instruments using children as content experts. *Nurs Res* 2005;54:414–8.
- [65] Tomlinson D, Gibson F, Treister N, et al. Understandability, content validity, and overall acceptability of the Children's International Mucositis Evaluation Scale (ChIMES): child and parent reporting. *J Pediatr Hematol Oncol* 2009;31:416–23.
- [66] Schilling LS, Dixon JK, Knaf KA, et al. Determining content validity of a self-report instrument for adolescents using a heterogeneous expert panel. *Nurs Res* 2007;56:361–6.
- [67] Harris SR, Daniels LE. Content validity of the Harris Infant Neuromotor Test. *Phys Ther* 1996;76:727–37.
- [68] Palisano RJ, Rosenbaum P, Bartlett D, et al. Content validity of the expanded and revised Gross Motor Function Classification System. *Dev Med Child Neurol* 2008;50:744–50.
- [69] Cox MJ, Paley B. Families as systems. *Annu Rev Psychol* 1997;48:243–67.
- [70] Pleil AM, Kimel M, McCormack J, et al. Psychometric Assessment of the Injection Pen Assessment Questionnaire (IPAQ): measuring ease of use and preference with injection pens for human growth hormone. *Health Qual Life Outcomes* 2012;10:126.
- [71] Ungar WJ, Mirabelli C, Cousins M, et al. A qualitative analysis of a dyad approach to health-related quality of life measurement in children with asthma. *Soc Sci Med* 2006;63:2354–66.
- [72] Ungar WJ, Boydell K, Dell S, et al. A parent-child dyad approach to the assessment of health status and health-related quality of life in children with asthma. *Pharmacoeconomics* 2012;30:697–712.
- [73] Picchietti DL, Arbuckle RA, Abetz L, et al. Pediatric restless legs syndrome: analysis of symptom descriptions and drawings. *J Child Neurol* 2011;26:1365–76.
- [74] Knaf K, Deatrick J, Gallo A, et al. The analysis and interpretation of cognitive interviews for instrument development. *Res Nurs Health* 2007;30:224–34.
- [75] Varni JW, Curtis BH, Abetz LN, et al. Content validity of the PedsQL 3.2 Diabetes Module in newly diagnosed patients with type 1 diabetes mellitus ages 8–45 [published online ahead of print December 27, 2012]. *Qual Life Res*. <http://dx.doi.org/10.1007/s11136-012-0339-8>.
- [76] Chang PC, Yeh CH. Agreement between child self-report and parent proxy-report to evaluate quality of life in children with cancer. *Psychooncology* 2005;14:125–34.
- [77] 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 4.0 Generic Core Scales. *Health Qual Life Outcomes* 2007;5:2.
- [78] Landgraf J, Abetz L, Ware J. *The CHQ User's Manual*. (2nd ed.). Boston: Health Act, 1999.
- [79] Riley AW, Forrest CB, Starfield B, et al. The Parent Report Form of the CHIP-Child Edition: reliability and validity. *Med Care* 2004;42:210–20.
- [80] Varni JW, Seid M, Kurtin PS. Pediatric health-related quality of life measurement technology: a guide for health care decision makers. *JCOM* 1999;6:33–40.
- [81] Pickard AS, Lin HW, Knight SJ, et al. Proxy assessment of health-related quality of life in African American and white respondents with prostate cancer: perspective matters. *Med Care* 2009;47:176–83.
- [82] Beattie PE, Lewis-Jones MS. A comparative study of impairment of quality of life in children with skin disease and children with other chronic childhood diseases. *Br J Dermatol* 2006;155:145–51.
- [83] Brown GT, Wright FV, Lang BA, et al. Clinical responsiveness of self-report functional assessment measures for children with juvenile idiopathic arthritis undergoing intraarticular corticosteroid injections. *Arthritis Rheum* 2005;53:897–904.
- [84] Brunner HI, Klein-Gitelman MS, Miller MJ, et al. Health of children with chronic arthritis: relationship of different measures and the quality of parent proxy reporting. *Arthritis Rheum* 2004;51:763–73.
- [85] Hutchings HA, Upton P, Cheung WY, et al. Development of a parent version of the Manchester-Minneapolis quality of life survey for use by parents and carers of UK children: MMQL-UK (PF). *Health Qual Life Outcomes* 2008;6:19.
- [86] Cremeens J, Eiser C, Blades M. Factors influencing agreement between child self-report and parent proxy-reports on the Pediatric Quality of Life Inventory 4.0 (PedsQL) generic core scales. *Health Qual Life Outcomes* 2006;4:58.
- [87] Davis E, Nicolas C, Waters E, et al. Parent-proxy and child self-reported health-related quality of life: using qualitative methods to explain the discordance. *Qual Life Res* 2007;16:863–71.
- [88] De Los Reyes A, Kazdin AE. Measuring informant discrepancies in clinical child research. *Psychol Assess* 2004;16:330–4.
- [89] Jozefiak T, Larsson B, Wichstrom L, et al. Quality of life as reported by school children and their parents: a cross-sectional survey. *Health Qual Life Outcomes* 2008;6:34.
- [90] Kiss E, Kapornai K, Baji I, et al. Assessing quality of life: mother-child agreement in depressed and non-depressed Hungarian. *Eur Child Adolesc Psychiatry* 2009;18:265–73.
- [91] Vance YH, Morse RC, Jenney ME, et al. Issues in measuring quality of life in childhood cancer: measures, proxies, and parental mental health. *J Child Psychol Psychiatry* 2001;42:661–7.
- [92] Zhou H, Roberts P, Horgan L. Association between self-report pain ratings of child and parent, child and nurse and parent and nurse dyads: meta-analysis. *J Adv Nurs* 2008;63:334–42.
- [93] Upton P, Lawford J, Eiser C. Parent-child agreement across child health-related quality of life instruments: a review of the literature. *Qual Life Res* 2008;17:895–913.
- [94] Canning EH, Hanser SB, Shade KA, et al. Maternal distress and discrepancy in reports of psychopathology in chronically ill children. *Psychosomatics* 1993;34:506–11.
- [95] Guyatt GH, Juniper EF, Griffith LE, et al. Children and adult perceptions of childhood asthma. *Pediatrics* 1997;99:165–8.
- [96] Levi RB, Drotar D. Health-related quality of life in childhood cancer: discrepancy in parent-child reports. *Int J Cancer Suppl* 1999;12:58–64.
- [97] Food and Drug Administration. *Guidance for Industry—Acute Bacterial Otitis Media: Developing Drugs for Treatment*. Silver Spring, MD: Food and Drug Administration, 2012.
- [98] Janse AJ, Sinnema G, Uiterwaal CS, et al. Quality of life in chronic illness: children, parents and paediatricians have different, but stable perceptions. *Acta Paediatr* 2008;97:1118–24.
- [99] Klassen AF, Miller A, Fine S. Agreement between parent and child report of quality of life in children with attention-deficit/hyperactivity disorder. *Child Care Health Dev* 2006;32:397–406.
- [100] Keller VE, Keck JF. An instrument for observational assessment of nausea in young children. *Pediatr Nurs* 2006;32:420–6.
- [101] Kleinman L, Rothman M, Strauss R, et al. The infant gastroesophageal reflux questionnaire revised: development and validation as an evaluative instrument. *Clin Gastroenterol Hepatol* 2006;4:588–96.
- [102] Annett RD. Assessment of health status and quality of life outcomes for children with asthma. *J Allergy Clin Immunol* 2001;107(Suppl.): S473–81.
- [103] Broder HL, McGrath C, Cisneros GJ. Questionnaire development: face validity and item impact testing of the Child Oral Health Impact Profile. *Community Dent Oral Epidemiol* 2007;35(Suppl. 1):8–19.
- [104] van Laerhoven H, van der Zaag-Loonen HJ, Derkx BH. A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatr* 2004;93:830–5.
- [105] Christie MJ, French D, Sowden A, et al. Development of child-centered disease-specific questionnaires for living with asthma. *Psychosom Med* 1993;55:541–8.
- [106] Flesch R. A new readability yardstick. *J Appl Psychol* 1948;32:221–33.
- [107] Prout HT, Chizik R. Readability of child and adolescent self-report measures. *J Consult Clin Psychol* 1988;56:152–4.
- [108] Likert RA. A technique for the development of attitude scales. *Educ Psychol Meas* 1952;12:313–5.
- [109] Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. New York: Oxford University Press, 2008.

- [110] Shields BJ, Palermo TM, Powers JD, et al. Predictors of a child's ability to use a visual analogue scale. *Child Care Health Dev* 2003;29:281–90.
- [111] Herdman M, Rajmil L, Ravens-Sieberer U, et al. Expert consensus in the development of a European health-related quality of life measure for children and adolescents: a Delphi study. *Acta Paediatr* 2002;91:1385–90.
- [112] Riley AW, Forrest CB, Rebok GW, et al. The Child Report Form of the CHIP-Child Edition: reliability and validity. *Med Care* 2004;42:221–31.
- [113] Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Med Care* 2001;39:800–12.
- [114] Baxter SD, Guinn CH, Royer JA, et al. Accuracy of children's school-breakfast reports and school-lunch reports (in 24-h dietary recalls) differs by retention interval. *Eur J Clin Nutr* 2009;63:1394–403.
- [115] Juniper EF, Guyatt GH, Feeny DH, et al. Measuring quality of life in the parents of children with asthma. *Qual Life Res* 1996;5:27–34.
- [116] Holme SA, Man I, Sharpe JL, et al. The Children's Dermatology Life Quality Index: validation of the cartoon version. *Br J Dermatol* 2003;148:285–90.
- [117] Tomlinson D, Gibson F, Treister N, et al. Refinement of the Children's International Mucositis Evaluation Scale (ChIMES): child and parent perspectives on understandability, content validity and acceptability. *Eur J Oncol Nurs* 2010;14:29–41.
- [118] Chambers CT, Craig KD. An intrusive impact of anchors in children's faces pain scales. *Pain* 1998;78:27–37.
- [119] French DJ, Christie MJ, Sowden AJ. The reproducibility of the Childhood Asthma Questionnaires: measures of quality of life for children with asthma aged 4–16 years. *Qual Life Res* 1994;3:215–24.
- [120] Valla JP, Bergeron L, Berube H, et al. A structured pictorial questionnaire to assess DSM-III-R-based diagnoses in children (6–11 years): development, validity, and reliability. *J Abnorm Child Psychol* 1994;22:403–23.
- [121] Whitcomb SA, Merrell KW. Interviewing techniques. In: *Behavioral, Social, and Emotional Assessment of Children and Adolescents* (4th ed.). New York: Routledge, 2013: chapter 6.
- [122] Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health* 2009;12:419–29.
- [123] Mangunkusumo RT, Duisterhout JS, de Graaff N, et al. Internet versus paper mode of health and health behavior questionnaires in elementary schools: asthma and fruit as examples. *J Sch Health* 2006;76:80–6.
- [124] Mussaffi H, Omer R, Prais D, et al. Computerised paediatric asthma quality of life questionnaires in routine care. *Arch Dis Child* 2007;92:678–82.
- [125] Raat H, Mangunkusumo RT, Landgraf JM, et al. Feasibility, reliability, and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF): internet administration compared with the standard paper version. *Qual Life Res* 2007;16:675–85.
- [126] Varni JW, Limbers CA, Burwinkle TM, et al. The ePedsQL in type 1 and type 2 diabetes: feasibility, reliability, and validity of the Pediatric Quality of Life Inventory Internet administration. *Diabetes Care* 2008;31:672–7.
- [127] 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.
- [128] Stritzke WG, Dandy J, Durkin K, et al. Use of interactive voice response (IVR) technology in health research with children. *Behav Res Methods* 2005;37:119–26.
- [129] Bullinger M, Alonso J, Apolone G, et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. *International Quality of Life Assessment*. *J Clin Epidemiol* 1998;51:913–23.
- [130] 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 for Translation and Cultural Adaptation. *Value Health* 2005;8:94–104.
- [131] Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46:1417–32.
- [132] Smythe I, Everatt J, Al-Menaye N, et al. Predictors of word-level literacy amongst grade 3 children in five diverse languages. *Dyslexia* 2008;14:170–87.
- [133] Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171–8.
- [134] Leidy NK, Revicki DA, Geneste B. Recommendations for evaluating the validity of quality of life claims for labeling and promotion. *Value Health* 1999;2:113–27.
- [135] Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–9.
- [136] Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. *Pediatrics* 2005;115:e600–14.
- [137] Prosser LA, Hammitt JK, Keren R. Measuring health preferences for use in cost-utility and cost-benefit analyses of interventions in children: theoretical and methodological considerations. *Pharmacoeconomics* 2007;25:713–26.
- [138] Ungar WJ. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated? *Pharmacoeconomics* 2011;29:641–52.
- [139] Wright DR, Wittenberg E, Swan JS, et al. Methods for measuring temporary health states for cost-utility analyses. *Pharmacoeconomics* 2009;27:713–23.
- [140] Kromm SK, Bethell J, Kraglund F, et al. Characteristics and quality of pediatric cost-utility analyses. *Qual Life Res* 2012;21:1315–25.
- [141] Ratcliffe J, Flynn T, Terlich F, et al. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. *Pharmacoeconomics* 2012;30:713–27.
- [142] Ravens-Sieberer U, Wille N, Badia X, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res* 2010;19:887–97.
- [143] Stevens K. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics* 2012;30:729–47.
- [144] Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res* 2010;19:875–86.
- [145] Prosser LA. Comparative effectiveness and child health. *Pharmacoeconomics* 2012;30:637–45.
- [146] Shaddy RE, Denne SC. The Committee on Drugs and Committee on Pediatric Research. Clinical report—guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations. *Pediatrics* 2010;125:850–60.
- [147] Shilling V, Williamson PR, Hickey H, et al. Processes in recruitment to randomised controlled trials of medicines for children (RECRUIT): a qualitative study. *Health Technol Assess* 2011;15:1–116.