NIH Public Access
Author Manuscript
Expert Rev Pharmacoecon Outcomes Res. Author manuscript; available in PMC 2011 November 1.

Published in final edited form as:
Expert Rev Pharmacoecon Outcomes Res. 2010 August ; 10(4): 385–396. doi:10.1586/erp.10.52.

Conceptual and methodological advances in child-reported outcomes measurement

Katherine B Bevans[†,1], **Anne W Riley**[2], **JeanHee Moon**[3], and **Christopher B Forrest**[1]

[1]Children's Hospital of Philadelphia and University of Pennsylvania School of Medicine, Department of Pediatrics, PA, USA

[2]Johns Hopkins Bloomberg School of Public Health, MD, USA

[3]Children's Hospital of Philadelphia, PA, USA

Abstract

Increasingly, clinical, pharmaceutical and translational research studies use patient-reported outcomes as primary and secondary end points. Obtaining this type of information from children themselves is now possible, but effective assessment requires developmentally sensitive conceptual models of child health and an appreciation for the rapid change in children's cognitive capacities. To overcome these barriers, outcomes researchers have capitalized on innovations in modern measurement theory, qualitative methods for instrument development and new computerized technologies to create reliable and valid methods for obtaining self-reported health data among 8–17-year-old children. This article provides a developmentally focused framework for selecting child-report health assessment instruments. Several generic health-related quality of life instruments and the assessment tools developed by the NIH-sponsored Patient-Reported Outcome Measurement Information System network are discussed to exemplify advances in the measurement of children's self-reported health, illness, wellbeing and quality of life.

Keywords

child; child health; classical test theory; development; item response; parent report; proxy report; psychometric; self-report; subjective well-being; theory

Patient-reported outcomes (PROs) are now common end points in clinical, pharmaceutical and translational research. Multiple initiatives are underway to ensure that a solid scientific foundation exists for the use of PRO assessments in clinical research and care. The US FDA recently issued a set of methodological standards that PRO assessment instruments must meet if they are used to formulate claims for medication and medical devices [1]. The NIH supports several trans-NIH initiatives that are developing a large number of health and wellbeing PRO measurement tools that are beginning to be used in research. These include the Patient Reported Outcome Measurement Information System (PROMIS) [2,3], NIH Toolbox [4] and Neuro-QOL [5]. These and other similar initiatives support the development, selection and employment of psychometrically robust PRO measurement

© 2010 Expert Reviews Ltd

[†]Author for correspondence: bevans@email.chop.edu.

Financial & competing interests disclosure The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.
No writing assistance was utilized in the production of this manuscript.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

tools, which are increasingly considered a core competencies of health services, outcomes and comparative effectiveness researchers [6].

In addition to these general research trends, it is increasingly understood that many precursors to adult health and disease are evident in childhood (e.g., early-onset symptoms, health-related behavior and exposure to health-compromising stressors) [7]. The potential to improve child health and prevent or lessen the impact of detrimental and costly chronic diseases that arise in adulthood through the early identification of precursor health states provides additional motivation for obtaining children's perspectives on their health. However, the assessment of child-reported health outcomes is a relatively understudied topic that faces a multitude of theoretical and methodological challenges [8-10]. Among the most significant challenges associated with the assessment of child PROs are the lack of consensus on the conceptualization of child health, illness, wellbeing and quality of life, and imperfect knowledge regarding how these outcomes are expressed among children at different developmental levels. PRO measures that are developed for adults are unlikely to capture the realities of childhood or to be sensitive to developmental change. Therefore, a thorough understanding of the developmental processes that influence children's social, cognitive and biologic competencies and challenges must be incorporated into the conceptualization of child PROs and, ultimately, their measurement [11].

The purpose of this article is to integrate the scientific and practical knowledge from multiple disciplines that support the valid and feasible assessment of child PROs, which include children's perspectives on their own health, illness, wellbeing and quality of life experiences. A framework for selecting child PRO instruments and methods is proposed here. The framework takes into account the current conceptual models of child health and illness outcomes, limitations in children's cognitive capacities and understanding of health and illness concepts, advanced methods for evaluating the psychometric properties of instruments, and innovative assessment methodologies. Although there are many factors that impact the children's capacity to provide reliable and accurate accounts of their own health status (e.g., emotional state and social desirability), we focus on the cognitive aspects including general cognitive competencies and understanding of health/illness concepts because awareness of children's rapidly expanding cognitive abilities is of particular importance when selecting PRO measurement tools for use with children. The article concludes with a summary of advances in pediatric PRO measurement that can be expected in the next 5 years.

Cognitive development & child self-report

An assumption underlying PRO measures is that individuals themselves are the most reliable and accurate observers of their health perceptions and experiences [9]. Limitations in general cognitive competencies, self-awareness, and understanding of health and illness concepts pose barriers to self-reporting among children [8,12]. A model developed by Demetriou and colleagues, which draws on information processing, psychometric and developmental theories to describe the architecture and development of the human mind highlights the many facets of cognitive functioning that may influence the reliability and accuracy of child-reported health [13,14]. According to this model, the mind is organized into three levels: processing potentials, domain-specific systems of thought and hypercognition. Processing potentials are the information processing mechanisms that underlie the ability to attend to, select, represent and act on information. They include speed of processing (the maximum speed at which a given mental act can be efficiently executed), control of processing (executive functions that enable an individual to attend to goal-relevant information while filtering out goal-irrelevant information and inhibiting premature responses) and representational capacity (working memory functions that support the mental

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

representation of information and the capacity to operate on it). Strengths and limitations in processing potentials, which vary across the course of child development, influence the functioning of domain-specific thought (systems that specialize in the representation/processing of specific types of information including categorical, quantitative, causal, spatial, propositional and social thought) and hypercognition (self-regulatory strategies used to achieve mental and behavioral goals).

Table 1 presents descriptions of how the three levels of cognitive function vary by age and influence the reliability and validity of children's self-reported health. Suggestions are provided for specific assessment strategies that capitalize on children's cognitive strengths and accommodate for their cognitive limitations at different developmental stages.

Children's understanding of health & illness concepts

As with more general cognitive and linguistic competencies, the development of children's understanding of illness occurs sequentially and progressively [15,16]. With age, children gain knowledge of a broader array of illness terminology, while the depth of their understanding increases as well. Young children (3–7 years of age) often explain illness in terms of proximal, concrete and sometimes magical causal phenomena (e.g., 'people get colds from the sun'). Older children (8–11 years old) tend to attribute illness to learned causes that are most often physical and external to the individual (e.g., germs cause illness; the sun causes sunburn). Most adolescents (12 years or older) are able to distinguish between the various causes of different diseases and recognize that illnesses are often initiated through interacting events such as infection combined with low levels of immunity [16].

In contrast to illness, relatively less is known regarding children's understanding of health concepts. There is evidence that even very young children can provide sophisticated definitions of health that recognize both biomedical (e.g., symptoms, life expectancy) and holistic (e.g., wellbeing, happiness and social support) components [17,18]. In their definitions of health, children as young as 8 years of age have been found to reference wellbeing, psychosocial health, participation in health-promoting behavior and avoidance of health-compromising factors [18,19]. Some developmental trends in children's self-reported health have been noted. In general, younger children report that they feel healthier and engage in fewer health-promoting behaviors than older children [20]. In addition, compared with adolescents, preadolescent children tend to believe that risky behaviors such as smoking, drinking alcohol and illicit drug use have greater potential to negatively influence health [21]. During middle childhood, awareness emerges that positive health behaviors such as adequate physical activity, nutrition, personal hygiene and sleep are essential for maintaining health and preventing the onset of disease [18]. Also, by preadolescence most children understand that environmental factors such as exposure to the sun and second-hand smoke can influence health [18,19].

Criteria for selecting child PRO measurement tools

In the past decade, there has been a rapid expansion in the number of PRO instruments that are specifically designed for and validated among children [22]. As with PRO assessment among adults, the measurement of pediatric PROs has been greatly improved by increased attention to the content of assessment tools (e.g., ensuring that content is related to the way that treatment is expected to work; thinking about research claims that the PRO will support) and the psychometric properties of tools as evidenced by both qualitative (e.g., ensuring that the measure taps the full range of patients' experiences) and quantitative (e.g., subjecting instruments to rigorous statistical tests) methods [23]. When selecting a PRO instrument, it is important to consider whether the tool suits the purpose of the investigation and if the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

dimensions covered are relevant to the respondents' context and experiences. It is also essential to evaluate the tool's sensitivity to developmental differences, psychometric properties in terms of reliability and validity, sensitivity to change (if the aim is to evaluate the effectiveness of an intervention or monitor changes in health status over time), and utility in clinical research or care [22]. Criteria for assessing the properties of child PRO assessment tools are reviewed below and summarized in Table 2.

For illustrative purposes, this article is focused on the development and validation of several prominent generic child-report health-related quality of life instruments as well as the outcome instruments generated by the PROMIS network [3]. PROMIS is a NIH effort that is changing the way patient-reported health outcomes are assessed and used in clinical research [2]. PROMIS has produced numerous psychometrically robust item banks for the assessment of self-reported physical, emotional and social health using modern test theory and information technology [3].

Developmental context of child PROs

Valid and clinically relevant measures of child PROs need to be based on a conceptual framework that accounts for children's development [10]. Virtually all well validated child PRO instruments have been designed to ask questions in language and formats that children can understand. This is accomplished by ensuring that questions are written well and can be commonly read by children aged 8–9 years old, sometimes accompanied by illustrations or size-graduated response options to aid in understanding, and tested for understanding in cognitive interviews. However, it must be said that PRO instruments developed for use with children of a wide age range are likely to be relatively insensitive to age-linked experiences such as those commonly observed during normative transitions into adolescence or young adulthood. In part, this is because many questions that are relevant and understood by youth of one developmental level are not appropriate for youth of another developmental level, and therefore cannot be retained in a fixed-format scale that is used with individuals of a broad age range. For example, in developing the Children's Health and Illness Profile (CHIP), several questions regarding trust and dependability in peer relationships were found to be reliable, valid and useful for the assessment of health among adolescents [24] but these items had inadequate psychometric properties when administered to children [25]. The concept of 'trust' was not reliably interpreted among preadolescent children. When the CHIP was revised (now called the Healthy Pathways Scales), the child and adolescent versions were combined to allow for the assessment of health among youth during the childhood-to-adolescence transition [26]. Among other revisions to enhance the developmental appropriateness of the tool, new items were generated and validated to capture the concept of trust in the peer relationships of youth, regardless of their developmental level (e.g., 'How often do you talk about everything with your friends?').

Qualitative methods such as small group interviews, focus groups, and cognitive interviews are valuable methods for tailoring item wording, item format and presentation so they are optimally understood by respondents. Such research is critical for obtaining children's perspectives on how they experience outcomes in various contexts (e.g., home, school or with peers), and to reveal the thought patterns and vocabulary that children commonly use in relation to PROs [12,27,28]. For example, focus groups conducted with children and their parents to assess the developmental appropriateness of the PROMIS definitions of PROs and the wording of items was essential to identifying gaps in coverage of the PROMIS PRO domains, as well as keywords, phrases and/or quotes that describe child PRO experiences [27,28].

Once items are developed, cognitive interviews are important to ascertain children's comprehension of PRO questions (e.g., what does the respondent believe the question is

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

asking? What do specific words and phrases mean to the respondent?); the process through which the child retrieves information from memory (e.g., what does the respondent do/need to recall the PRO information?); the decision-making process (e.g., is the answer influenced by motivation to respond or social desirability?); and the response process (e.g., can the respondent match his/her answer to the response categories?) [29]. Focus groups, cognitive interviews and other qualitative research methods ensure that children's voices are heard and respected in the instrument development process and often result in significant modifications to instrument content (e.g., item deletion, addition and/or rewording) [12,27,28].

Once the child health assessment questionnaire or other assessment tool is complete, it is administered to a large number of respondents, yielding data that are analyzed using classical and modern methods to assess the dimensionality of the PROs assessed, the consistency with which they measure specific concepts (reliability), and the degree to which they measure a full range of the concepts they purport to measure (validity).

Reliability

Internal consistency reliabilities for most commonly-used child PRO assessment instruments are acceptable with expected variation across subpopulations [30]. For example, internal consistency reliabilities of the Healthy Pathways Scales are good to excellent (α range: 0.75–0.86), with a definite age gradient such that younger children are less reliable, although still acceptable reporters [8]. Likewise, each child-reported scale of the Child Health Questionnaire has moderate to excellent internal consistency (range: 0.67–0.89) and for six of the ten scales, all constituent items are correlated with their parent construct at a level of 0.40 or higher [31]. Test–retest reliability remains to be assessed for most child PRO instruments. However, for those evaluated, there is moderate to strong stability in children's reports over time [22]. For example, all domains of the CHIP – Child Edition (CHIP-CE) demonstrate good to very good stability over a 2-week period (intra-class correlation: 0.63–0.76) [25].

Validity

An instrument's criterion validity is assessed by evaluating associations between children's scores on the instrument and independent measures of similar constructs. For several instruments, scores on child PRO instruments are consistent with relevant externally derived information. The CHIP-CE criterion validity is demonstrated through significant relations between children's domain scores and scores on measures of internalizing emotions, physical functioning, activity limitations, social competence, academic performance and self-esteem [25]. Similarly, high and moderate correlations were found between psychosomatic complaints and the KIDSCREEN moods and emotions, psychological wellbeing, and self-perception dimensions [32]. The KIDSCREEN-52 [33] and the CHIP-CE [25] were useful in predicting the frequency of visits to health professionals and hospitalization. Finally, children's self-reports of school performance on the CHIP – Adolescent Edition (CHIP-AE) [24] and the Pediatric Quality of Life Inventory (PedsQL™) [34] are consistent with objective indictors of academic achievement.

Another way to assess measurement validity is to compare scores among groups of children that are expected to differ on PROs. These comparisons have been made for child PRO instruments on the basis of known stressors (e.g., illness, low socioeconomic status and maltreatment), gender and age. For some PRO assessment tools, construct validity is supported by significant relations between child-reported outcomes and the presence of a chronic health condition. For example, adolescents with asthma and vision problems report poor perceived health on the Child Health Questionnaire compared with peers without significant health concerns [35]. Similarly, the PedsQL™ is effective at distinguishing

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

between healthy children and those with chronic health conditions [36]. Other confirming relations have been noted between child-reported health and social stressors such as family poverty [32,37].

Modern measurement approaches

Classical test theory (CTT) methods such as those described earlier, are based on the assumption that respondents' observed scores are an estimate of their true scores plus or minus some unobservable measurement error. CTT methods depend on few *a priori* assumptions and produce results that are relatively easy to interpret. However, these methods are limited in that true scores are not exclusively based on respondent characteristics, but may also depend on the content of the assessment tool [38,39]. Therefore, it cannot be determined whether variability in scores is attributable to respondents' outcome state or the properties of the PRO items administered. As a result, it is difficult to compare the outcomes of respondents based on results of different health assessment instruments [40]. Also, CTT methods assume (often erroneously) that the reliability of a measure remains constant for all respondents regardless of their level on the health construct being assessed and that the relationships among items and between items and the underlying traits they are purported to measure are linear in nature [38].

By contrast, a unique aspect of item response theory (IRT) is that it supports the differentiation of respondent and item characteristics through the property of parameter invariance [39]. A cornerstone of all IRT models is the joint measurement of people and items, specifically that a respondent's 'ability' (level of the underlying PRO dimension) and item 'difficulty' (how much of the outcome a person must have to endorse the item) are placed on the same continuum and assessed using the same metric. The joint measurement of respondent and item parameters ensures that IRT procedures are uniquely positioned to:

- Provide information regarding gaps in the measurement of the underlying construct (e.g., when there are no or too few items that provide information regarding respondents along the full range of the PRO continuum) and item redundancy (e.g., when too many items provide information regarding respondents at a specific point on the PRO continuum);
- Provide a natural framework for studying item bias through the comparison of varying subgroup members with comparable levels of health on the probability of endorsing specific items [39];
- Support the development of computerized adaptive test (CAT) versions of instruments that 'match' the most appropriate set of items to each respondent based on their level of underlying PRO [41,42]. In CAT assessment, the sequence of items administered depends on an individual's response to previously administered items. Based on the respondent's prior performance, items that provide the most information regarding their unique level of the PRO are administered. In this way, a small number of items could be applied without sacrificing measurement precision. Likewise, it is possible to develop short instruments that still enable precise measurement.

Increasingly, child PRO assessment instruments are developed and validated using modern measurement approaches. For example, both the Healthy Pathways Scales [26] and the KIDSCREEN [32] were developed and validated using a combination of classical and modern (i.e., IRT) psychometric techniques. In addition, PROMIS provides an unparalleled model for using modern psychometric methods to construct, analyze and refine item banks that assess PROs for both children and adults [2]. Specifically, PROMIS instruments are systematically developed using advanced factor analytic methods and IRT approaches to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

select items that are both comprehensive and efficient in assessing health outcomes, remove items that are biased against key demographic and clinical groups, and generate and test CAT algorithms [2,3].

Innovative methods of assessing child-reported health

Over the past decade, there has been a proliferation of computer-assisted and other multimedia methods in health research, which reflects a general trend toward leveraging technology to improve the accuracy and efficiency of data capture procedures. For survey research in particular, digital technology has helped to address numerous methodologic and cost-efficiency challenges. Child health assessment measures have been enhanced with graphical, video and audio content in an effort to improve children's understanding of health terms, concepts and engagement in the assessment process [8,26]. The presentation of PRO survey items in pictorial formats that depict specific symptoms, attributes, skills and activities has been shown to improve younger children's engagement in the assessment process [25,43] and the reliability of their responses [44]. For example, a computer-based version of the child self-report Strengths and Difficulties Questionnaire presents simple colorful illustrations of age, gender and race-neutral figures that illustrate item content. Compared with traditional paper-and-pencil version of the Strengths and Difficulties Questionnaire, the computerized version had greater test–retest reliability and parent–child concordance, as well as higher ratings of child satisfaction and engagement in the assessment process [45].

In addition to illustrating item content, illustrated response options are sometimes used to enhance children's capacity to accurately answer survey items. Visual analog scales provide a reliable method for children to communicate about the frequency or intensity of health-related experiences such as pain or other physical and emotional experiences sensations and feelings [46]. In completing the Violence Exposure Scale for Children-Revised [47] children aged 6–8 years select one out of four thermometers to indicate the frequency with which they have been exposed to violence. This measure has been shown to produce reliable data, even for children in preschool and kindergarten. In developing the CHIP, Rebok and colleagues determined that children aged 6–11 years preferred response categories that represent greater frequency or intensity with increasingly large 'graduated' circles [12].

Another technique that accommodates for children's limited attentional capacities as well as their developing reading and linguistic abilities is the collection of health-related information via audio computer-assisted self interview (ACASI), a computer application that allows a research participant to hear survey interview items over a computer headset and read the corresponding items on a computer monitor. ACASI automates progression from one item to the next and can be programed to skip irrelevant items. Interviewees respond to questions by pressing a number keypad, using a touch screen or clicking on a response, thereby sending the data directly into a database [48]. Compared with other data collection methods including paper self-administered questionnaires, telephone interviews and in-person interviews, ACASI reduces under-reporting bias among adults and adolescents for sensitive information such as same-gender sexual contact and substance abuse [49,50]. Furthermore, ACASI supports application of CAT in which interviewees' prior responses elicit the presentation of the most relevant follow-up questions [51].

Technology also supports the use of innovative time-dependent data collection methods such as momentary time sampling, a method in which responses are recorded at predetermined intervals (e.g., an individual's rating of pain every 30 min) [52]. Unlike retrospective methods, momentary time sampling procedures do not depend on the ability to summarize events that occurred in the distant past. Historically, momentary time sampling data have

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

been collected using paper diaries; however, the feasibility and accuracy of these methods have been improved with the availability of personal digital assistants, text messaging and cell phones [53]. These technologies provide a means to remind research participants to record data and a device through which data can be recorded, stored and electronically transmitted to a secure database. Recent research suggests that up to 90% of paper diary respondents forge data (e.g., enter multiple backlogged entries instead of completing the diary in real time) and electronic diaries contain significantly fewer completion errors [52]. Interestingly, boys demonstrate greater compliance with electronic diary formats than girls, highlighting the need to consider different technology platforms for use with various population subgroups [54].

Looking forward, there are many other technological innovations that may improve the assessment of children's health, particularly when used in combination with PRO measurement tools. For example, gaming consoles, such as the Nintendo® Wii™, have been used to augment physical rehabilitation in children and adults. The Wii is based on the concept of virtual reality, an immersive, interactive, 3D computer experience occurring in real time. The Wii's signature feature is the motion-sensitive remote controller, which allows an individual to interact with and manipulate items on screen via gesture recognition and pointing through the use of accelerometer and optical sensor technology. Most simply, the use of a gaming console like Wii or gaming joystick represents yet another vehicle to more actively engage children, through a medium that is both fun to use and already familiar to many of them [55]. Further research should systematically and intensively test innovative data collection techniques that may be used in combination with children's self-report because both methods provide unique and complementary information.

Assessment of child health through parent report

Despite compelling rationale for measuring child perspectives on their health and illness outcomes, considerable evidence of children's reliability and accuracy as self-reporters, and the many advances in child-reported health instruments and methodologies, assessment of child PROs through proxy report is still advisable under some circumstances. For example, primary caregivers (usually parents) may be asked to complete PRO assessment tools on behalf of their very young or developmentally delayed children because they are presumed to have intimate knowledge of their children's health, illness, wellbeing, quality of life and service utilization. When children are able to report on their own behalf, parents may provide a unique and complementary perspective on their children's outcomes, but their reports should not be used to discount the views of children themselves [8,10,56]. Unfortunately, the often observed low rates of parent–child concordance on outcome measures are frequently used to support the erroneous assumption that parents are always more accurate and reliable reporters than are children themselves [56,57]. A number of factors affect parent–child agreement including the child's age, the PROs assessed, child health status and the quality of parent–child relationships [56,58]. Future research should further elucidate how and under what conditions various reporters provide the most accurate assessments of child PROs. Such research will lead to improved methods for selecting the best reporter (e.g., child vs parent) or combining multiple reports to enhance measurement reliability, accuracy and clinical validity.

Expert commentary

The increasing prioritization of PROs in clinical, pharmaceutical and translational research has introduced the patient's perspective into understanding the effects of clinical interventions. Developmentally appropriate and psychometrically robust instruments for the assessment of child PROs are increasingly available. Improvement in the measurement of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

child PROs are occurring in the large part, because of the integrated efforts of experts from many disciplines including pediatrics, public health, developmental and cognitive psychology, psychometrics, child computer interaction and health informatics. The integration of these diverse fields of study presents a challenge, but also incredible opportunities for rapid innovation. For example, modern measurement approaches (e.g., IRT) have been combined with advances in computer technology, providing a powerful framework for administering individually tailored PRO assessments via CAT. These methods have been shown to improve measurement precision, reduce response burden and identify bias in the measurement of PROs [2]. Recent advances in the measurement of child PROs have occurred in large part because of the multiple initiatives to ensure that a solid scientific foundation exists for the use of PRO assessments in pediatric clinical research and care (e.g., PROMIS, the FDA's methodological standards for the assessment of PROs).

Five-year view

It is our belief that over the next 5 years, the rapid advances in PRO measurement will continue with a particular focus on the assessment of PROs among children. These efforts will be enhanced by psychometric innovations that will support the more common use of multidimensional IRT and test equating (e.g., across multiple forms or respondents), and increased precision in CAT. The measurement of PROs will be enhanced by increased sensitivity to the diverse and rapidly changing developmental competencies of people across the life-course. Increasingly, there will be attempts to 'match' innovative data collection methods to the cognitive/linguistic strengths and limitations characteristic of each developmental stage. Finally, with increased recognition of PROs as essential end points in clinical, pharmaceutical and translational research, the development and adoption of methodological standards for the development and validation of PRO instruments will be emphasized.

Acknowledgments

Anne W Riley manages the CHIP instrument and the royalties collected for commercial use of the CHIP.

References

Papers of special note have been highlighted as:

- • of interest
- •• of considerable interest

1. United States Department of Health & Human Services Food and Drug Administration. Guidance for industry patient report outcome measures: use in medical product development to support labeling claims. March 15.2010

2•. Reeve B, Hays RD, Bjorner J, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcome Measurement Information System (PROMIS). Med Care. 2007; 45(5):S22–S31. Provides an overview of the methods used to assess the psychometric properties of Patient-Reported Outcomes Measurement Information System assessment tools. [PubMed: 17443115]

3. National Institutes of Health. Patient-Reported Outcomes Measurement Information System: dynamic tools to measure health outcomes from the patient perspective. Mar 15.2010

4. Gershon RC, Cella D, Fox NA. Assessment of neurological and behavioural function: the NIH Toolbox. Lancet Neurol. 2010; 9(2):138–139. [PubMed: 20129161]

5. Neuro-QOL: Quality of life in neurological disorders. Jun.2010

6. Forrest CB, Martin DP, Holve E, Millman A. Health services research doctoral core competencies. BMC Health Serv Res. 2009; 9:107–111. [PubMed: 19555485]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

7. Forrest CB, Riley AW. Childhood origins of adult health: a basis for life-course health policy. Health Aff (Millwood). 2004; 23(5):155–164. [PubMed: 15371381]
8. Bevans, K.; Forrest, CB. The reliability and validity of children's self-reported health. In: Ungar, W., editor. Economic Evaluation of Child Health. Oxford; NY, USA: 2010.
9. DeCivita M, Reiger D, Alamgir AH, Anis AH, FitzGerald MJ, Marra CA. Evaluating health-related quality-of-life studies in paediatric populations: some conceptual, methodological and developmental considerations and recent applications. Pharmacoeconomics. 2005; 23(7):659–685. [PubMed: 15987225]
10. Ravens-Sieberer U, Erhart M, Wille N, Wetzel R, Nickel J, Bullinger M. Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. Pharmacoeconomics. 2006; 24(12):1199–1220. [PubMed: 17129075]
11. Forrest CB, Simpson L, Clancy C. Child health services research: challenges and opportunities. JAMA. 1997; 22:1787–1793. [PubMed: 9178792]
12. Rebok G, Riley A, Forrest C, Starfield B, Green B, Robertson J. Elementary school aged children's reports of their health: a cognitive interviewing study. Qual Life Res. 2001; 10(1):59–70. [PubMed: 11508476]
13. Demetriou A, Raftopoulos A. Modeling the developing mind: from structure to change. Dev Rev. 1999; 19:319–368.
14. Demetriou, A.; Efklides, A. Structure, development, and dynamics of the mind: a meta-Piagetian theory. In: Demetriou, A.; Efklides, A., editors. Mind, intelligence, and Reasoning: Structure and Development. Elsevier; Amsterdam, The Netherlands: 1994. p. 75-109.
15. Williams JM, Binnie LM. Children's concepts of illness: an intervention to improve knowledge. Br J Health Psychol. 2002; 7(2):129–147. [PubMed: 14596705]
16. Solomon GE, Cassimatis NL. On facts and conceptual systems: young children's intergration of their understandings of germs and contagion. Dev Psychol. 1999; 35(1):113–126. [PubMed: 9923469]
17. Piko B. Health-related predictors of self-perceived health in a student population: the importance of physical activity. J Community Health. 2000; 25(2):125–137. [PubMed: 10794206]
18. Piko BF, Bak J. Children's perceptions of health and illness: images and lay concepts in preadolescence. Health Educ Res. 2006; 21(5):643–653. [PubMed: 16740672]
19. Oakley A, Bendelow G, Barnes J, Buchanan M, Husain OA. Health and cancer prevention: knowledge and beliefs of children and young people. BMJ. 1995; 310(6986):1029–1033. [PubMed: 7728055]
20. Woods SE, Springett J, Porcellato L, Dugdill L. 'Stop it, it's bad for you and me': experiences of and views on passive smoking among primary-school children living in Liverpool. Health Educ Res. 2005; 20(6):645–655. [PubMed: 15829496]
21. Piko B. Smoking in adolescence: do attitudes matter? Addict Behav. 2001; 26(2):201–217. [PubMed: 11316377]
22••. 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. 2007; 11(4):742–764. Describes the content and psychometric properties of currently available health-related quality of life instruments for use with children and adolescents. [PubMed: 18179668]
23. Shiffman S. Two decades of change for PROs: how patient-reported data collection has been transformed since 1987. Appl Clin Trials. 2008; 17(1):66–66.
24. Starfield B, Riley A, Breen B, Ensminger M, Ryan S, Kelleher K. The adolescent child health and illness profile: a population-based measure of health. Med Care. 1995; 33(5):553–566. [PubMed: 7739277]
25. Riley AW, Forrest CB, Rebok GW, Starfield B, Green BF, Robertson JA. The child report form of the CHIP-Child edition: reliability and validity. Med Care. 2004; 42(3):221–231. [PubMed: 15076821]
26. Bevans KB, Riley AW, Forrest CB. Development of the healthy pathways child-report scales. Qual Life Res. In press.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

27. Irwin DE, Varni JW, Yeatts K, DeWalt DA. 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. [PubMed: 19166601]
28. Walsh TR, Irwin DE, Meier A, Varni JW, DeWalt DA. The use of focus groups in the development of the PROMIS pediatrics item bank. Qual Life Red. 2008; 17:725–735.
29. Willis, GB. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Sage; CA, USA: 2005.
30. 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(1):37–45. [PubMed: 14706404]
31. Landgraf, JM.; Abetz, L.; Ware, JE. Child Health Questionnaire (CHQ): A User's Manual. Vol. 1. The Health Institute, New England Medical Center; MA, USA: 1996.
32. Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Duer W. KIDSCREEN-52 quality-of-life measure for children and adolescents. Expert Rev Pharmacoecon Outcomes Res. 2005; 5(3): 353–364. [PubMed: 19807604]
33. Rajmil L, Alonso J, Berra S, Ravens-Sieberer U, Gosch A, Simeoni M. Use of a children's questionnaire of health related quality of life (KIDSCREEN) as a measure of need for health care services. J Adolesc Health. 2006; 38(5):511–518. [PubMed: 16635761]
34. Varni JW, Burwinkle TM, Seid M. The PedsQL (TM) 4.0 as a school population health measure: feasibility, reliability, and validity. Qual Life Res. 2006; 15(2):203–215. [PubMed: 16468077]
35. Waters EB, Salmon LA, Wake M, Wright M, Hesketh KD. A school-based population study of the self-report child health questionnaire. J Adolesc Health. 2001; 29:140–149. [PubMed: 11472873]
36. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory Version 4.0 Generic Core Scales in healthy and patient populations. Med Care. 2001; 39(8):800–812. [PubMed: 11468499]
37. von Rueden U, Gosch A, Rajmil L, Bisegger C, Ravens-Sieberer U. Socioeconomic determinants of health related quality of life in childhood and adolescence: results from a European study. J Epidemiol Community Health. 2006; 60(2):130–135. [PubMed: 16415261]
38. van der Linden, WJ.; Hambleton, RK. Handbook of Modern Item Response Theory. Springer-Verlag; NY, USA: 1997. Item response theory: brief history, common models, and extensions; p. 1-28.Anonymous
39. Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. Educ Measurement Issues and Practice. 1993:38–47.
40•. Hambleton RK. Emergence of item response modeling in instrument development and data analysis. Med Care. 2000; 38(9):II60–II65. Highlights advances in modern measurement theory and its application to health outcomes measurement. [PubMed: 10982090]
41. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. Qual Life Res. 2007; 16(Suppl. 1):95–108. [PubMed: 17530450]
42••. Cella D, Gershon R, Lai J, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res. 2007; 16:133–141. Describes methods of generating and validating item banks and computerized adaptive tests to assess patient-reported outcomes. [PubMed: 17401637]
43. Harter S, Pike R. The pictorial scale of perceive competence and social acceptance for young children. Child Dev. 1984; 55(6):1969–1982. [PubMed: 6525886]
44. Valla JP, Bergeron L, Bidaut-Russell M, St-George M, Gaudet N. Reliability of the Dominic-R: a young child mental health questionnaire combining visual and auditory stimuli. J Child Psychol Psyc. 1997; 38(6):717–724.
45. Truman J, Robinson K, Evans AL, et al. The Strengths and Difficulties Questionnaire: a pilot study of a new computer version of the self-report scale. Eur Child Adoles Psy. 2003; 12(1):9–14.
46. McGrath PA, Seifert CE, Speechley KN, Booth JC, Stitt L, Gibson MC. A new analogue scale for assessing children's pain: an initial validation study. Pain. 1996; 64(3):435–443. [PubMed: 8783307]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

47. Shahinfar A, Fox NA, Leavitt LA. Preschool children's exposure to violence: relation of behavior problems to parent and child reports. Am J Orthopsychiatry. 2000; 1(115):125.
48. Jones R. Survey data collection using audio computer assisted self-interview. West J Nurs Res. 2003; 25(3):349–358. [PubMed: 12705116]
49. Simoes AA, Bastos FI, Moreira RI, Lynch KG, Metzger DS. Acceptability of audio computer-assisted self-interview (ACASI) among substance abusers seeking treatment in Rio de Janeiro, Brazil. Drug Alcohol Depend. 2006; 82(Suppl. 1):S103–S107. [PubMed: 16769438]
50. Villarroel MA, Turner CF, Eggleston E. Same-gender sex in the United States: impact of the T-ACASI on prevalence estimates. Public Opin Q. 2006; 70(2):166–196. [PubMed: 21998488]
51. Wainer, H. Computerized Adaptive Testing: A Primer. Lawrence Erlbaum Associates; NJ, USA: 2000.
52. Stone AA, Broderick JE, Schwartz JE, Sfiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. Pain. 2003; 104(1–2):343–351. [PubMed: 12855344]
53. Dale O, Hagen KB. Despite technical problems personal digital assistants outperform pen and paper when collecting patient diary data. J Clin Epidemiol. 2006; 60(1):8–17. [PubMed: 17161749]
54. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain. 2004; 107(3):213–219. [PubMed: 14736583]
55. Deutsch JE, Borbely M, Filler J, Huhn K, Guarrera-Bowlby P. Use of a low-cost, commercially available gaming console (Wii) for rehabilitation of an adolescent with cerebral palsy. Phys Ther. 2008; 88(10):1196–1207. [PubMed: 18689607]
56. 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(6):895–913. [PubMed: 18521721]
57. Barker ET, Bornstein MH, Putnick DL, Hendricks C, Suwalsky JT. Adolescent-mother agreement about adolescent problem behaviors: direction and predictors of disagreement. J Youth Adolesc. 2007; 36:950–962.
58. Davis E, Nicolas C, Waters E, et al. Parent-proxy and child self-report health-related quality of life: using qualitative methods to explain the discordance. Qual Life Res. 2007; 16:863–871. [PubMed: 17351822]
59. Bland JM, Altman DG. Statistics notes: Cronbach's α. BMJ. 1997; 314(7077):572–574. [PubMed: 9055718]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Key issues

- Increasingly, clinical, pharmaceutical and translational research studies rely on patient-reported outcomes (PROs) as primary and secondary end points.
- An assumption underlying PRO measures is that individuals themselves are reliable and accurate observers of their health perceptions and experiences.
- Strengths and limitations in general cognitive competencies, self-awareness and understanding of health and illness concepts impact the reliability and accuracy of child PROs.
- Child PRO measurement tools should be based on a conceptual framework that accounts for children's development and explicitly tested for children's understanding of health/illness-related concepts and terminology.
- Classical and modern measurement approaches are applied to assess the dimensionality of PRO measurement tools, the consistency with which they measure specific concepts (reliability), and the degree to which they measure a full range of the concepts they purport to measure (validity).
- Psychometric methods rooted in item response theory are useful for identifying gaps/redundancies in the measurement of PROs and item bias, and support the development of computerized adaptive test versions of PRO instruments.
- Increasingly, computer-assisted and other multimedia data collection methods are being used to improve the accuracy and efficiency of PRO measurement.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Cognitive capabilities, understanding of health and illness concepts, and suggested assessment strategies for children of varying developmental levels.

Cognitive capabilities (processing potentials)	Impact on health reporting	Age range		
		3–5 years	*6–11 years*	*12–17 years*
Processing speed	Impacts efficiency in processing questions, retrieving relevant information from memory, integrating information into a summative judgment, interpreting and choosing response options, and communicating the final answer	Processing speed is considerably slower than that of older children; can interfere with children's understanding of health-related questions, retrieval of relevant information from memory and reasoning/problem solving	Significant increases in processing speed, but limitations are still evident. Children aged 8–10 years typically respond to a range of motor, perceptual and cognitive tasks at speeds, that is, 0.5–0.6 standard deviations below that of young adults	Processing speed continues to improve systematically with age, but children aged 12–13 years still respond at a speed more than a full standard deviation below that of young adults. Processing speed approaches ceiling during young adulthood and is maintained until middle age
Control of processing	Impacts capacity to focus on relevant information while filtering out irrelevant information (both in interpreting questions and generating responses), suppress irrelevant or impulsive/ premature responses to questions, and identify/ implement a plan for responding to items	By age 3, children have developed rudimentary attentional control and are increasingly able to switch rapidly between two simple stimuli. Simple planning skills are exhibited by many 4-year-olds, but younger children are generally unable to plan and organize actions in advance	Attentional control continues to improve with age, and children aged 9 years or older tend to monitor and regulate their actions well and are able switch rapidly between tasks. Cognitive planning and organizational skills develop rapidly between 7–10 years of age	Sustained attention, behavioral inhibition and switching fluency continue to improve during adolescence. Problem solving and decision making strategies are significantly refined
Representational capacity	The capacity to temporarily store and process information impacts the comprehension of questions, reasoning required to select the most appropriate response, and evaluating the answer (which may include editing owing to social desirability or context)	By age 3, children can formulate mental representations of objects that are not present, but mental manipulation of these objects is limited	Representation capacity improves dramatically during middle childhood, but the capacity to store and manipulate information is still limited, particularly for complex information. Limitations in representational capacity may compromise processing speed	The capacity to temporarily store and process information improves significantly during adolescence. Beginning at approximately 12 years of age, adolescents are able to mentally represent and manipulate increasingly abstract information [13]
Domain-specific systems of thought	Domain-specific systems of thought specialize in the representation/ processing of specific types of information (e.g., categorical, quantitative, causal, spatial, propositional and social). Errors in the representation or processing of information increases the likelihood of misinterpreting or misunderstanding health-related questions. Errors in processing increases the likelihood of inaccurate responding	Errors in the selection and use of cognitive operations are somewhat common, which may result in misinterpretation of questions and inaccurate responses	Increasingly, children avoid errors in the selection and use of domain-specific operations but their capacity to engage in domain-specific thought is not equally developed across domains (e.g., a child may be more proficient at quantitative than spatial thought). The reliability and accuracy of child-reported health may be enhanced through the use of assessment methodologies that present information through multiple modalities (e.g., present questions in written and	Adolescents may operate on domain-specific information using the computational or operational rules of another system. For example, they may encode and process using graphical representations (part of the spatial domain) to express categorical relationships (part of the categorical domain). This emerging capacity improves encoding, processing, and retrieval of information

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Cognitive capabilities (processing potentials)	Impact on health reporting	Age range		
		3–5 years	*6–11 years*	*12–17 years*
			oral formats simultaneously)	
Hypercognition	Impacts the capacity to attend to assessment tasks, allocate the mental resources necessary to accurately interpret and answer questions, select the most appropriate problem-solving operation and modify responses by evaluating problem-solving outcomes	Young children have limited hypercognitive capabilities and, therefore, are generally unable to understand the demands to evaluate their experiences or to regulate their own problem solving and decision-making skills	Children have self-awareness capacities; however, communication between the hypercognitive and domain-specific systems is somewhat limited. Children tend not to purposefully self-monitor or self-regulate their own problem solving and decision making	The childhood-to-adolescence transition is marked by improved communication between the hypercogntive and domain-specific systems; adolescents monitor and modify their attention, behavior and problem solving with increased efficiency
Understanding of health and illness concepts	Impacts children's capacity to understand a question and generate a reliable and accurate response	Understanding of health and illness centers around occurrence of specific symptoms (e.g., earache). Minimal understanding of health/illness terminology. Children attribute health/illness to concrete, proximal and often infeasible phenomena	Developing understanding of health/illness terms and phrases. Substantial increase in children's understanding of health/illness. Tendency to attribute health/illness to concrete environmental conditions or behaviors (e.g., smoking)	Substantially improved understanding of health/illness terms and phrases. Understanding that health/illness is attributable to complex interactions between self and environment, which can change over time. Concept of health expands to include wellbeing, social connectedness and functioning
Suggested assessment strategies		Obtaining caregiver/parent report of child health is generally required for a complete assessment of health, but it is recommended that children's perspectives also be obtained. One-on-one interviews may help children to maintain attention and understand the assessment content	Assessment tools must be tested to ensure that items are understandable to children. Illustrated items or response categories may improve interest, attention and engagement in the assessment process. Items should be read aloud to children or administered via ACASI format to compensate for diversity in reading capabilities. Adequate time allocation needed for completion	Many (although not all) adolescents can complete self-administered health questionnaires that are administered in written format (e.g., a paper and pencil questionnaire) without assistance; Interest, engagement and attentional capacities can be improved through the use of engaging assessment formats (e.g., ACASI), but many adolescents dislike illustrated items

ACASI: Audio computer-assisted self interview.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Methods of assessing the developmental appropriateness, dimensionality and psychometric properties of child-reported outcome instruments.

Consideration	Definition	Methods
Conceptual model		
Developmental sensitivity of the model	Valid and clinically relevant measures of pediatric PROs are based on a conceptual framework that accounts for children's development	Review the instrument's underlying conceptual framework to determine if it is consistent with empirically supported developmental theory; qualitative methods (e.g., focus groups and cognitive interviews) are often used to refine the conceptual model based on children's reported health experiences and to ensure that specific items are understandable and meaningful to children
Dimensionality	Summary of the theoretical and empirical relations between a set of items resulting in qualitatively distinct (i.e., unidimensional) constructs or 'factors'	Exploratory and confirmatory factor analyses or principal component analyses are used to model observed variables (e.g., responses to items) as linear combinations of factors plus 'error' terms. Each subscale should include only a single factor
Reliability		
Internal consistency	Assesses how strongly related three or more items that purport to measure the same construct are to one another	Usually measured with Cronbach's α, a statistic calculated from the pairwise correlations between items. Commonly accepted rule: α of 0.7 or higher indicated acceptable to good reliability [59]. Extremely high reliabilities (>0.95) are not necessarily desirable, as this indicates that the items may be entirely redundant
Test–retest reliability	Stability of the measure (the observed score) over time, assuming no change has occurred in underlying health	Typically, the same instrument is administered to the same respondents twice, within a few day interval; adequate reliability is indicated if the Pearson correlation between scores is high (~0.7 or higher) and if the Cohen's κ statistic is 0.6 or higher. Retest reliability provides more information regarding accuracy than other types of reliability
Inter-rater reliability	Agreement, homogeneity or consensus among independent raters answering on the same instrument (e.g., children and their parents)	Different statistics are appropriate for different types of measurement: Cohen's κ (for categorical items), values at 0.6 or better are acceptable; concordance correlation coefficient (for continuous data), value of 0.65 are acceptable; ICC (operates on data structured as groups, e.g., child and parent), ICCs of 0.6 or higher are acceptable
Validity		
Content	The degree to which all aspects of the underlying health dimension are included in a scale purporting to measure that health construct. In modern measurement terms, this would be extent to which all facets of the latent trait are included	Assurances of content validity are based on a prior conceptualization of the health construct assessed, and evaluation of the conceptualization and measurement tool by lay (i.e., structured interviews or focus groups) and expert judges. Qualitative research with potential respondents may be used to help define the nature of the construct
Construct	The 'summary' type of validity that reflects the extent to which the instrument measures the health construct it purports to measure. Subtypes: concurrent (relating a measure to other concrete criteria assessed simultaneously); predictive (the degree to which measure can predict future or independent past events)	Demonstrating construct validity is an ongoing process that depends on various validity tests. Concurrent validity is supported when correlations between scores on the health instrument and similar scales or variables that should be related are moderate/high and in the predicted direction. Predictive validity is demonstrated when scores are useful in predicting future outcomes
Modern measurement approaches		
Coverage of dimension	Extent to which an instrument measures the full range of the underlying health dimension so that it accurately measures that aspect of health among all children no matter what their level of health is	IRT models are used to calibrate items based on an at minimum, the item 'difficulty'. Items with positive difficulty are 'harder' (e.g., endorsed only by individuals with high levels of the health latent trait) and items with negative difficulty are 'easier' (i.e., endorsed by many individuals with at least a minimum level of the health trait). The 'distance' between items on a scale in terms of their difficulty should not

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Consideration	Definition	Methods
		be so large that meaningful points along the continuum of the health dimension are not adequately measured by the scale
Item uniqueness/redundancy	Extent to which each item on a health assessment instrument provides unique information regarding children with a specific level of the underlying health dimension	Inspection of item difficulty as described earlier is used to ensure that only a single or a small number of items on a single scale are equally difficult. It is generally desirable for item difficulties to be unique within each scale to ensure comprehensive measurement of the underlying health construct with as few items as possible
Item bias	An item is biased by group membership (e.g., age, gender) if individuals with the same level of health, but from different groups answer the item in a systematically different way; that is, they have different probabilities of endorsing the item	IRT is used to detect DIF, which is indicated when an item is more 'difficult' for one subgroup than another or if the item is more useful at discriminating among children with varying levels of health within one subgroup compared with another. For example, item DIF is evidenced when members of different subgroups (e.g., boys and girls) who have the same level of the underlying health construct have different probabilities of endorsing the item

DIF: Differential item functioning; ICC: Intra-class correlation; IRT: Item response theory; PRO: Patient-reported outcome.