Quality of Life Research (2005) 14: 1045–1056
DOI 10.1007/s11136-004-4714-y

Proxy assessment of quality of life in pediatric clinical trials: Application of the Health Utilities Index 3

Cheryl L. Cox[1], Shelly Lensing[2], Shesh N. Rai[2], Pam Hinds[1], Elizabeth Burghen[1], & Ching-Hon Pui[3]
[1]*Department of Nursing Research (E-mail: cheryl.cox@stjude.org);* [2]*Department of Biostatistics;* [3]*Department of Hematology-Oncology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA*

Accepted in revised form 5 October 2004

Abstract

Background: With increased cure rates, pediatric oncology protocols increasingly seek to document the impact of treatment on patients' disease, symptoms, and functional capacity. *Procedure*: Nurses as proxy respondents used the Health Utilities Index 3 (HUI3) to assess the health-related quality of life (HRQL) in twenty-five patients (age 6 years or older) enrolled on a frontline protocol for leukemia. HRQL observations (n = 70) were made at three different time points to coincide with high-dose methotrexate therapy. Additionally, the proxy respondents evaluated the ease of use of the instrument and the data quality. *Results*: As patients' health status declined, the number of unassessable HRQL items increased. These missing data made scoring cumbersome and precluded calculation of the overall HRQL scores for nearly 50% of the patients. *Conclusions*: Use of the provider proxy-assessed HUI3 in pediatric cancer trials may result in a high proportion of missing data. Trials may benefit more from the use of HRQL measures that consider the acuity of the child's illness, domains specific and sensitive to both disease and treatment, and items that can be proxy-assessed independent of input from parent or patient. Evaluations that combine child self-reports with both parent and provider reports may ultimately provide the most reliable and comprehensive perspective on children's quality of life.

Key words: Clinical trial, Health-related quality of life, Pediatric oncology

Introduction

Pediatric clinical trials have begun to address the impact of therapy on patients' health-related quality of life (HRQL) as well as assessing the standard end points of response rate, probability of survival, and toxicity. The inclusion of HRQL assessment in clinical trials addresses the need for: (1) holistic evaluation of interventions; (2) incorporation into subsequent treatment protocols of patient and family perceptions of response to treatment; (3) a mechanism to facilitate decisions about individual patient treatment options; and (4) patient-centered data that extend beyond clinical measures of efficacy to better assess the cost-effectiveness of treatment [1, 2].

Assessment of HRQL is less straightforward in pediatric patients than in adults, and methods are less well developed [3–5]. While a variety of instruments that target children's HRQL are available, the optimal indicators of HRQL remain subject to debate [3–7]. Some investigators feel strongly that it is the subjective, not objective, reality of the patient's health status that is relevant to HRQL; however, others see the potential for large 'discrepancies' between subjective and objective evaluations of the impact of disease and treatment on health status, and that it is the objective data that should ultimately describe health status [8–9]. Selected HRQL measures have therefore emerged that reject personal judgments and feelings as the only source of information [10–11]. Multi-attribute utility measures, or health

status classification systems, weight a subject's perceptions about his or her current health status (as measured across several instrument-determined functional attributes), according to predominant popular perceptions about different health states. The result is an overall HRQL (health utility) measure.

Here we report our experience with the use of a proxy-administered health utilities instrument to measure the HRQL of children enrolled in a clinical oncology trial. Specifically, we defined HRQL as a multidimensional construct which describes subjective perceptions of health status; these perceptions are influenced by disease and its treatment, and they are assessable on selective physical, functional, perceptual, and social parameters. We viewed HRQL as a concept which has meaning to and is understood by the general public and health professionals [12]. Our complete design to assess HRQL in children in the clinical trial included a planned comparison of child self-report, parent report, and proxy report. Because of the likely toxic effects of the treatment protocol and the associated high rate of self-report missing data during times of high acuity, we sought a proxy-respondent measure that would allow an HRQL evaluation that relies on observation independent of child or parent self-report. The Health Utilities Index 3 [13] represented a multi-dimensional HRQL construct, available in a provider proxy assessable format, and weighted in accordance with previously determined social valuations of health states. Our results demonstrate that the use of this instrument to measure children's HRQL entails conceptual and methodological problems. We offer recommendations to improve the assessment of pediatric patients' HRQL in clinical trials.

Background

The symptomatic sequelae of diseases and experimental therapies have prompted the development of research protocols that are patient-centered rather than strictly disease-oriented [14, 15]. Patients are increasingly given choices among physician-defined therapeutic options [16], and the patient's HRQL is a significant factor in these choices. In pediatric oncology, HRQL is an increasingly important medical goal as the likelihood of survival increases. Today's clinical research protocols are charged with documenting not only the objective physiological impact of treatment (disease response and toxicity) but also the patients' perception of the impact of the treatment on their lives (symptoms, physical/mental/social function, and life satisfaction).

There has been extensive debate in the pediatric health outcomes literature about the validity of proxy-respondents in evaluating children's HRQL [5, 7, 17]. There is clear support for obtaining HRQL information directly from the child; however, when the child is too young or too ill to provide the information directly, parents and/or caregivers have served as proxy respondents. The findings are inconsistent with respect to the agreement between proxy and child reports on HRQL domains [5, 7]. Higher agreement between child and parent is associated with domains that are more directly observable–physical activity, functioning, and symptoms; lower agreement between parent and child is associated with non-observable emotional domains [7, 17]. A similar type of correspondence between health professionals' and children's ratings of HRQL domains is reported; agreement is higher on scales measuring somatic distress and activity than on scales measuring more emotional or internal states [7]. In proxy assessment of health utility in children, reliability and interpretation have varied with the source of information and the modality of assessment (see Table 2) [18–20].

Despite inconsistent findings about the congruency of proxy and self-report measures of HRQL, proxy measures may be the only available approach in some cases. In longitudinal clinical trials, for example, patients' conditions may be compromised by their disease, treatment, or both, and their ability to directly respond to HRQL assessments is often diminished. In such cases, proxy respondents can offer distinct methodological advantages [21]. Through the use of an HRQL instrument that lends itself to direct observation, the frequent problem of missing data in clinical trials potentially can be overcome or lessened. Additionally, the biases attributed to parent reports (overestimation of the impact of disease and treatment on the child) [5, 17, 22] may be avoided. In this context we sought to gather HRQL directly from the child through qualitative interviews, and

from the parent(s) through self-report; to address the potential for missing child data in times of high acuity and toxicity, and the potential for biased data from the parents, we sought to rely on proxy direct observation as an additional HRQL assessment.

Methods

We used the proxy assessment format of the HUI3 [13] to measure HRQL in children enrolled on a frontline institutional treatment protocol designed to improve the cure rate and quality of life of children and adolescents with acute lymphoblastic leukemia. All study procedures were in accordance with the institutional review board and the US-DHHS Office for Human Research Protections. The risk-directed treatment protocol included a number of primary and secondary (pharmacologic, infectious disease, biologic, and cancer control) research aims. The main phases of therapy were remission induction (6–7 weeks duration, with optional high-dose methotrexate in the up-front window), consolidation (2 weeks), and continuation (120 weeks for females and 146 weeks for males).

Study design

All newly enrolled patients on the treatment protocol who were 6 years of age or older and whose parents had given permission for their child to participate were eligible for the study. Nurses provided serial proxy HRQL observations by using the HUI3 at 4 time points: Week 6 of induction, Week 7 of continuation, Week 31 of continuation, and Week 120 or 146 of continuation (end of therapy). The first 3 data collection points were purposely scheduled to coincide with high-dose methotrexate therapy because of its' potential impact on symptoms and perceived functional health status. Data for these three time points are reported here. Because data for T_4 are presently being collected, these data will not be included in the analysis.

Instrument

The Health Utilities Index is a proprietary generic, multi-attribute QOL instrument [13]. Two complementary systems (HUI2 and HUI3) are available in a variety of formats (self-administered, interview, proxy assessment). Developers of the HUI systems define HRQL as proposed by Patrick and Erickson [23]: 'the value assigned to duration of life as modified by the impairments, functional states, perceptions, and social opportunities that are influenced by disease, injury, treatment, or policy.' HUI2, initially developed to assess outcomes among survivors of pediatric cancer [24], consists of 7 attributes or dimensions of health (sensation, mobility, emotion, cognition, self-care, pain, fertility). HUI3 [25], originally applied in the 1990 Statistics Canada Ontario Health Survey, consists of 8 attributes (vision, hearing, speech, ambulation, dexterity, emotion, cognition, pain). The HUI System has been applied to a wide variety of health issues in more than 25 countries and has been translated into more than 15 languages [11, 13]. Table 1 summarizes applications of the Health Utilities Index and Table 2 lists recent conceptual and methodological challenges encountered in using the HUI System.

Each of the 8 attributes is assessed at 5–6 levels of function ranging from normal to severely impaired function. In addition to a single 5-point (excellent to poor) Likert item to address overall health status (Question #41), the composite health status of a subject is described by combining the scores for 8 different attributes. Multiplicative multi-attribute utility functions (weights) translate the categorical data into interval-scale single-attribute utility scores (reflecting the morbidity in that dimension of health status); the single-attribute utility scores are then summed to produce an overall global QOL score, which describes the level of function in terms of the range 0 (dead) to 1.0 (perfect health). The health status measures offer a description of the type and extent of disability, while the utility scoring measures offer information about the relative importance of the disabilities from a social valuation perspective [13].

Depending on the focus of the study and the particular patient population, HUI2, HUI3, or both may be used. While the two instruments are roughly equivalent, they differ in concepts of emotion and pain; self-care and fertility assessment are available only in HUI2; and dexterity is assessed only in HUI3 [13, 26–28]. HUI3 was selected for this study because it offered the best fit

Table 1. Summary of Health Utilities Index System Applications, Reliability, and Validity

Evidence	Application	Reliability	Validity
24	Population Survey		Face validity
43	NICU survivors & matched random sample of school children		Construct validity
44	Children with Acute lymphocytic leukemia		Construct validity
45	Survivors of Acute lymphocytic leukemia		
27	Children with brain tumors	Inter-rater reliability between physicians, nurse, parents (Range $r = 0.57$–0.90)	
28	Survivors of childhood cancer		
24	Survivors of Acute lymphocytic leukemia		
46	Children admitted to the Intensive Care Unit	Inter-rater reliability $r > 0.80$	
47	Population Survey	Test–retest (moderate to better k coefficients for individual questions); overall $r = 0.767$	
48	Adult brain tumor vs. general population		Construct validity
49	Survivors of medulloblastoma		Concurrent validity
50	Survivors of childhood cancer		Concurrent validity
51	Survivors of Wilms tumor and neuroblastoma	↓ inter-rater reliability between parent and child on cognition	Construct Validity
52	Survivors of childhood cancer	Inter-rater agreement between parents and physicians	
29	Population survey		Predictive validity (intraclass correlation, $r = 0.88$)
53	Parents of children undergoing treatment for cancer vs. parents in general population		Construct validity (Intraclass correlation, $r = 0.99$)
54	Population survey		Construct Validity (Total potential error in estimates using scoring formula (SD = 0.06)

with our patient population. The dexterity assessment in the HUI3 was relevant to the treatment protocol, whereas neither the fertility nor the self-care domains in HUI2 were of immediate relevance to the treatment outcomes.

Setting and data collection

The study took place in a freestanding research institution whose mission is to find cures for catastrophic childhood illnesses. The institution has 56 inpatient beds and approximately 2000 outpatient contacts each week.

In the proxy-administered format of the HUI3, the proxy is to rely on observation, not parent or patient report. Nurses as proxy respondents used the HUI3 to assess HRQL 70 times across three data collection intervals; 20 assessments were completed by inpatient nurses and 50 were completed by outpatient nurses. Six of the 25 patients had all of their interviews completed by the same nurse, 6 had the same nurse for 2 of the three interviews, and 9 patients had a different nurse at each time point. Two patients had data collected at only two points by a different nurse each time. The mean time required to complete the instrument was 5.6 min by inpatient nurses and 4.2 min by outpatient nurses (range, 2–13 min).

The proxy respondents' qualitative assessments of the instrument's ease of administration and

Table 2. Conceptual and psychometric challenges in using Health Utilities Instruments

Authors	Challenges
References 55–57	Content validity: Health utilities are limited with respect to their standing as either a health status measure or quality of life measure. Important health status criteria (neuropsychological and psychosocial functioning) are absent; likewise, criteria that would define quality of life more broadly than perceived health status and functional performance are missing
References 18–20, 64	Reliability: In children, results of HUI3 vary with source of information and the modality of administration; estimates by patients and 'experts' differ; greatest agreement is between parents and patients, with some tendency for parents to underestimate health status while physicians overestimate; MDs and physiotherapists differ on emotion, cognition, pain; higher inter-rater and inter-modality agreement for physical attributes and poor agreement for psychological attributes.
References 58–63	Sensitivity/Specificity: Decreased sensitivity of HUI2/3 due to limited number of health aspects addressed and crude scale response options; cultural insensitivity; more a measure of disability than quality of life; responds to changes in health status associated with serious chronic illnesses with predominant alterations in sensation or cognition, but not consistent with changes in self-reported health. May not be appropriate for populations with severe disabilities since floor effects are unlikely given the scale's design
References 10, 61	Scoring: Summary scores across multiple attributes with an assigned utility may be useless in a clinical trial (summary score may show no difference between groups). Response options in utilities are limited and contribute to potential floor and ceiling effects.

their perceptions of the quality of the data obtained were prospectively built into the study. Specifically, the nurses were asked to address these questions after they had completed the HUI3 for the first time (within several minutes or up to 2 h after completing their first patient assessments).

Results

Patients

Twenty seven of the 53 patients enrolled on the treatment protocol met our study criteria. Fifteen of the 27 patients were female and 18 were white. Median age at the time of enrollment was 10 years (range, 6–18 years). At the time of this report, 21 (84%) patients were alive with a median of 3.8 years (range, 0.3-4.5 years) of follow-up. Quality of life data were not collected for 2 patients who died before the first assessment; therefore, the total sample size available for analysis at T_1 was 25. One patient was taken off protocol; one patient died; and one patient was not evaluated at T_3. Therefore, the sample sizes were 25, 23, and 22 at the first, second, and third time points, respectively.

Completeness of responses

A significant quantity of data was missing, despite the fact that proxy respondents had undergone extensive orientation to the HUI3, as recommended [29, 30]. Speech was the only HRQL category for which there were no missing data. Table 3 and Figure 1 identify the total number and percentage of missing data at each time point and within each category. At T_1, data were missing at the highest rates for vision (9/25; 36%), emotion (6/25; 24%), and cognition (6/25; 24%). A score is required for each of the functional QOL attributes, without exception, to allow calculation of the overall multi-attribute utility (MAU) score for QOL. Table 4 lists the summary statistics for MAU scores and for the eight HUI attributes. The MAU score could not be calculated for 60% (15/25) of patients at T_1, 44% (10/23) of patients at T_2, and 39% (9/22) of patients at T_3. Overall, we were unable to calculate 48.6% (34/70) of the MAU scores.

Problem HUI3 items

In QOL analysis, it is crucial to identify whether data are missing because of random or non-random factors [31, 32]. To address this question, we

Table 3. Distribution of number of missing attributes at each time point

No. of missing attributes per Time Point			Frequency (No. of patients)
T_1	T_2	T_3	
0	0	0	4
0	0	1	3
0	0	4	1
0	1	0	1
0	1	4	1
1	0	0	2
1	0	2	1
1	1	0	3
2	NA	NA	1
2	0	1	1
2	1	0	2
2	3	0	1
2	3	1	1
3	NA	NA	1
3	0	2	1
3	5	NA	1
Total number of patients:			25

NA: not assessed.

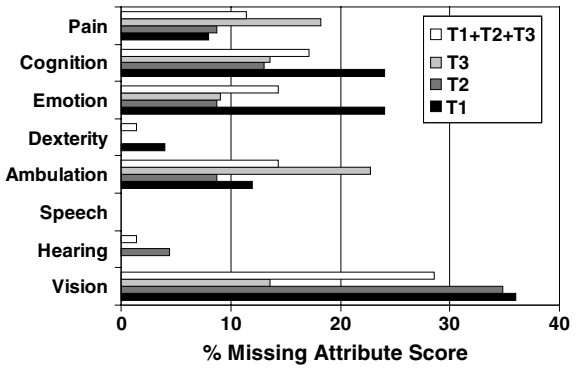

Figure 1. Percentage of missing scores for each HRQL attribute at each time point (n = 25, 23, and 22 at T_1, T_2, and T_3, respectively).

examined the attributes and specific items for which data were most often missing (Table 5).

A ceiling effect (i.e., most scores near the maximum value) was observed in selected categories (Table 6). For example, vision received a perfect score 49/50 times, hearing 69/69 times, and dexterity 67/69 times. Scores (number [score]) were slightly more varied for ambulation (51[1], 3[2], 2[5], 4[6]), emotion (41[1], 10[2], 7[3], 2[4]), cognition (51[1], 5[2], 1[3], 1[4]), and pain (34[1], 9[2], 9[3], 4[4], 6[5]).

Skip patterns in the HUI questions (responses that allowed some subsequent questions to be skipped) made it difficult to apply simple data imputation methods to compensate for missing data. One imputation method recommended by the HUI developers is hot-decking as described by Little and Rubin [33]. In this method, if imputation is at the level of the MAU score and the overall rate of missing data is not substantial, missing scores can be replaced with a score randomly selected from those of a similar individual. However, no recommendation is given as to the level of data at which hot-decking should be applied, i.e., MAU score, attribute score, or individual item [34]. In one application, the HUI developers replaced missing MAU scores with scores imputed from randomly selected respondents with similar response patterns on other questions [29]. Because of the high rate of missing data at assessments when overall health status was poor (as indicated by Question #41), this option was not deemed suitable. The lowest health status was reported for those patients for whom data were most often missing (Figure 2); paradoxically, quality of life assessment is of primary interest in these patients.

Proxy respondents' comments on HUI3

Many of the nurses wrote specific comments next to items that they were unable to assess. These notes indicated that 7 items were left blank because the nurses had no opportunity to observe the function in question (e.g., Can you see your friends across the street without glasses?) and 23 items were left blank because the child was too sedated or ill to demonstrate the level of function. Two nurses, in rating two different patients, noted that the observed impairment was not disease- or treatment-related, but rather represented a pre-existing condition (i.e., cerebral palsy). Thirteen nurses independently commented that the HUI3 appeared to measure functional performance rather than a child's perceived quality of life, and that they did not consider the assessment of function

Table 4. Summary statistics for MAU scores and attribute scores

Time	No. of Patients	Attribute	No. obs.[a]	Mean	Std Error	Median	Minimum	Maximum
MAU scores								
1	25		10	0.90	0.03	0.91	0.71	1.00
2	23		13	0.91	0.06	1.00	0.25	1.00
3	22		13	0.87	0.06	1.00	0.20	1.00
Attribute scores								
1	25	vision	16	1.3	0.3	1.0	1.0	6.0
		hearing	25	1.0	0.0	1.0	1.0	1.0
		speech	25	1.2	0.2	1.0	1.0	5.0
		ambulation	22	2.0	0.4	1.0	1.0	6.0
		dexterity	24	1.2	0.2	1.0	1.0	6.0
		emotion	19	1.6	0.2	1.0	1.0	4.0
		cognition	19	1.2	0.1	1.0	1.0	2.0
		pain	23	2.2	0.3	2.0	1.0	5.0
2	23	vision	15	1.0	0.0	1.0	1.0	1.0
		hearing	22	1.0	0.0	1.0	1.0	1.0
		speech	23	1.2	0.2	1.0	1.0	5.0
		ambulation	21	1.3	0.2	1.0	1.0	6.0
		dexterity	23	1.2	0.2	1.0	1.0	6.0
		emotion	21	1.3	0.2	1.0	1.0	4.0
		cognition	20	1.1	0.1	1.0	1.0	2.0
		pain	21	1.6	0.3	1.0	1.0	5.0
3	22	vision	19	1.0	0.0	1.0	1.0	1.0
		hearing	22	1.0	0.0	1.0	1.0	1.0
		speech	22	1.0	0.0	1.0	1.0	1.0
		ambulation	17	1.1	0.1	1.0	1.0	2.0
		dexterity	22	1.0	0.0	1.0	1.0	1.0
		emotion	20	1.6	0.2	1.0	1.0	3.0
		cognition	19	1.3	0.2	1.0	1.0	4.0
		pain	18	2.3	0.3	2.0	1.0	5.0

[a]Note smaller sample sizes due to missing items.

to be an adequately sensitive indicator of an ill child's experience of disease and treatment.

Developers of the HUI strongly recommend that preventive measures be taken to minimize missing data; they emphasize that each questionnaire should be thoroughly checked as soon as it is returned by the respondent '…and the respondent even called to resolve problems as soon as they are discovered…' [34]. These caveats seem most practical for self-reported or even parent-assessed health related quality of life, in that the data collector can identify missing items and ask for clarification. However, if the healthcare provider proxy respondent cannot address the items through observation, and must ultimately refer to the parent to complete the instrument, there appears to be little point in conducting a proxy HRQL assessment.

Discussion

First, we found that as overall health status declined (coinciding with higher toxicity), there was an increase in the number of HUI3 items that the nurses as proxy respondents were unable to assess. These findings are consistent with those obtained from self-reported HRQL instruments, which demonstrate higher proportions of 'don't know' and 'refused to answer' responses as treatment toxicity increases. It is well established, however, that a high proportion of missing data within and across time points precludes the accurate assessment of reliability, validity, and clinical significance [1].

Second, because we had prospectively selected a provider direct observation proxy format for our study deliberately to address the potential for

Table 5. HUI items that were missing data (n = 70 scores from 25 patients at 3 time points)

Question	N missing/N expected responses[a] (%)
Vision – 20/70 (28.6%) attributes not scored due to missing item response	
Have you been able to see well enough to recognize a friend on the other side of the street *with* glasses or contact lenses?	12/14 (85.7%)
Have you been able to see well enough to read ordinary newsprint *with* glasses or contact lenses?	7/13 (53.9%)
During the past four weeks, have you been able to see well enough to recognize a friend on the other side of the street *without* glasses or contact lenses?	14/69 (20.3%)
During the past four weeks, have you been able to see well enough to read ordinary newsprint *without* glasses or contact lenses?	10/70 (14.3%)
During the past four weeks, have you been able to see at all?	1/9 (11.1%)
Ambulation – 10/70 (14.3%) attributes not scored due to missing item response	
Have you needed the help of another person to walk?	2/7 (28.6%)
Have you been able to walk around the neighborhood *with difficulty* but *without help or equipment* of any kind?	4/14 (28.6%)
Have you been able to walk around the neighborhood *without difficulty* and *without help or equipment* of any kind?	4/22 (18.2%)
During the past four weeks, have you been able to bend, lift, jump and run *without difficulty* and *without help or equipment* of any kind?	8/70 (11.4%)
During the past four weeks, have you been able to walk at all?	1/11 (9.1%)
Have you needed mechanical support, such as braces or a cane or crutches, to be able to walk around the neighborhood?	0/7 (0.0%)
Have you needed a wheelchair to get around the neighborhood?	0/11 (0.0%)
Emotion – 10/70 (14.3%) attributes not scored due to missing item response	
Would you describe yourself as having felt: a) somewhat unhappy, b) very unhappy, c) so unhappy that life is not worthwhile	4/14 (28.6%)
During the past four weeks, have you been feeling happy or unhappy?	9/70 (12.9%)
Would you describe yourself as having felt: a) happy and interest in life, or b) somewhat happy?	4/60 (6.7%)
Cognition – 12/70 (17.1%) attributes not scored due to missing item response	
How would you describe your ability to remember things, during the past four weeks: a) able to remember most things, b) somewhat forgetful, c) very forgetful, d) unable to remember anything at all?	9/70 (12.9%)
How would you describe your ability to think and solve day to day problems, during the past four weeks: a) able to think clearly and solve problems, b) had a little difficulty c) had some difficulty, d) had a great deal of difficulty, e) unable to think or solve problems?	9/70 (12.9%)
Pain Attribute – 8/70 (11.4%) attributes not scored due to missing item response	
How many of your activities, during the past four weeks, were limited by pain or discomfort: none, a few, some, most, all?	8/36 (22.2%)
Have you had any trouble with pain or discomfort during the past four weeks?	3/70 (4.3%)

[a]Because of skip patterns (questions that could be skipped, depending on responses to prior questions), there were different denominators for each question.

missing or biased data, there was no means of retrieving or compensating for non-assessable items. Thus, overall MAU scores (health utility scores that reflect overall quality of life) could not be calculated for nearly 50% of our patients; planned comparisons between qualitative child reports, parent self-report and proxy reports could not be completed, and clinicians' ability to draw reasonable inferences about the impact of the disease and its' treatment was significantly compromised.

The HUI3 measures health status on the basis of functional capacity; however, our experience suggests that the functional attributes assessed by the HUI3 may be less relevant during treatment than after treatment for childhood cancer. Functional

Table 6. Percentage of patients with perfect (score = 1) attribute scores

Time	No. of Patients	Attributes	No. of Obs.[a]	Percent
1	25	vision	16	93.8
		hearing	25	100.0
		speech	25	92.0
		ambulation	22	77.3
		dexterity	24	95.8
		emotion	19	57.9
		cognition	19	84.2
		pain	23	43.5
2	23	vision	15	100.0
		hearing	22	100.0
		speech	23	95.7
		ambulation	21	90.5
		dexterity	23	95.7
		emotion	21	85.7
		cognition	20	95.0
		pain	21	76.2
3	22	vision	19	100.0
		hearing	22	100.0
		speech	22	100.0
		ambulation	17	88.2
		dexterity	22	100.0
		emotion	20	60.0
		cognition	19	84.2
		pain	18	44.4

[a]Note smaller sample sizes due to missing items.

attributes that might be more relevant to the effects of treatment either are not measured by HUI3 or are not measured at a useful level of discrimination. In this study, little to no variation was seen in speech, dexterity, vision, or hearing during treatment, except when attributable to some pre-existing condition. Greater variation among patients was observed in the categories of pain, cognition, and ambulation in this trial; however, an even broader dispersion of response options would have permitted a more discriminating assessment of change.

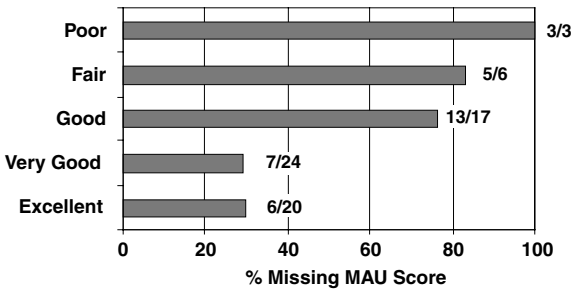

Figure 2. Percentage of missing multi-attribute utility scores according to health status across all 3 time points (n = 70 scores from 25 patients).

Despite intensive orientation and training in use of the HUI3, our experienced nurses encountered significant difficulty as proxy respondents. As toxic effects increased, categories of function could not be assessed on the basis of the available items. This observation is consistent with the diminished continuity of patient–clinician relationships and patient contact time as therapy is increasingly delivered in an outpatient setting. Therefore, the health-care professional proxy respondent may be increasingly unable to answer many of the items assessed by health utilities, and the resultant greater amount of missing or inaccurate data may cause large measurement errors. Other questions may not be answerable by the proxy respondent in any case, because they are not context-specific; for example, pediatric oncology nurses do not 'walk around the block' with their patients. Some other domains, while assessable, addressed areas of function that were unrelated to the cancer and its' treatment. The instrument does not allow separate evaluation of the impact of pre-existing conditions and therapy. If child self-report is not a viable option during the intensive treatment phase of a clinical trial, then assessable domains that lend themselves to direct observation by a proxy, and indicators that reflect a wider range of illness states and potential responses to toxic effects must be developed.

Conclusion

The use of a health utility measure in a nurse proxy-administered format in a pediatric oncology clinical trial is likely to be impeded by a large amount of missing data. In our case, the domains assessed were not sufficiently sensitive and specific for our purposes. The scoring schemes were cumbersome because of the missing data, and we were unable to measure subtle shifts in clinical status over time.

The measures of HRQL in clinical trials must be as robust as the measures used for other outcomes [35]. Trials that seek to document the impact of disease and its treatment are likely to benefit from using measures that are 'health status related' or 'disease and treatment specific.' At those time

points when the acute toxicity of therapy jeopardizes both self-report and proxy assessment, instruments more focused on observed signs, clinical indicators, altered responses, and changes in disposition [36] and on comparative (e.g., 'better or worse than yesterday') statements of health status might be optimal. Highly sensitive, expanded response options that facilitate the detection of subtle changes would also be appropriate. Condition-specific, disease-specific, and situation-specific instruments, like those developed in adult oncology [37], may prove to be more sensitive to subtle changes than the generic utility-based instruments. For example the *PedsQL* [38] consists of a core section that addresses physical, mental, and social health domains. To this core can be added disease-specific modules (cancer, diabetes, cystic fibrosis) [39].

Therapy trials for childhood cancer comprise multiple phases. One quality of life instrument may not adequately and sensitively address the subtleties of each phase. During early and late therapy, before toxic effects emerge and after they decline, truly subjective self-report measures that address the gap between the patient's immediate life situation and the patient's definition of a good quality of life are important and should be a legitimate endpoint in clinical trials. At such times, greater depth and specificity across additional domains (e.g., well-being, mental health, spiritual, existential) [30], or attributes that do not lend themselves as well to assessment during times of increased acuity, may be investigated (e.g., *Quality of Life Profile* [40], *How are You?* [41], *Comprehensive quality of life scale* [42]). The choice of instruments should reflect children's perception of their immediate quality of life beyond that which is strictly related to their health state; item content should be developmentally and contextually specific where appropriate.

When proxy assessment is required to meet study objectives and address data validity issues, the same methodological effort and rigor are required for proxy instrument development as for self-report instrument development to maximize content domain, context relevancy, and assessability. Respondent-tailored instruments that are specific to the proxy (parent or provider) should be developed in order to provide as broad a perspective as possible on the child's quality of life. Assessments that combine child self-reports with parent and provider assessments may ultimately provide the most reliable and comprehensive perspective on children's quality of life in clinical trials.

Acknowledgements

This work was supported in part by Grant CA21765, National Cancer Institute; American Cancer Society FM Kirby Clinical Research Professorship (C-H PUI), and by the American Lebanese Syrian Associated Charities (ALSAC).

References

1. Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health related quality of life research to support labeling and promotional claims in the United States. Qual Life Res 2000; 9(8): 887–900.
2. Scott D, Garrood T. Quality of life measures: Use and abuse. Baillere's Clinical Rheumatology 2000; 14(4): 663–687.
3. Wallander JL, Schmitt M, Koot HM. Quality of life measurement in children and adolescents: Issues, instruments, and applications. J Clin Psychol 2001; 57(4): 571–585.
4. Soni MK, Cella D. Quality of Life and symptom measures in oncology: An overview. Am J Managed Care 2002; 8(18 Suppl): s560–s573.
5. Matza LS, Swensen AR, Flood EM, Secnik K, Kline-Leidy N. Assessment of health-related quality of life in children: A review of conceptual, methodological, and regulatory issues. Value Health 2004; 7(1): 79–92.
6. Rajmil LR, Herdman M, Fernandez De Sanmamed M, et al. Generic health-related quality of life instruments in children and adolescents: A qualitative analysis of content. J Adol Health 2004; 34: 37–45.
7. Eiser C, Morse R. The measurement of quality of life in children: Past and future perspectives. Dev Behav Pediatr 2001: 22(4): 248–256.
8. Liang MH, Lew RA, Stucki G, Fortin PR, Daltroy L. Measuring clinically important changes with patient-oriented questionnaires. Med Care 2002; 40(4 Suppl): II45–II51.
9. Patrick DL, Chiang YP. Measurement of health outcomes in treatment effectiveness evaluations: Conceptual and methodological challenges. Med Care 2000; 38(9 Suppl): II14-II25.
10. Dijkers M. Measuring quality of life: Methodological issues. Am J Phy Med Rehabil 1999; 78: 286–300.
11. Horsman J, Furlong W, Feeny D, Torrance G. The health utilities index (HUI): Concepts, measurement properties and applications. Health and Quality of Life Outcomes 2003; 1: 54 (http://www.hqlo.com/content/1/1/54).

12. Fayers PM, Hand DJ, Bjordal K, Groenvold M. Causal indicators in quality of life research. Qual Life Res 1997;6(5): 393–406.
13. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The health utilities index (HUI) system for assessing health-related quality of life in clinical studies. Ann Med 2001; 33(5): 375–384.
14. Osoba D. What has been learned from measuring health related quality of life in clinical oncology? Euro J Cancer 1999; 35(11): 1565–1570.
15. Sanders C, Egger M, Donovan J, et al. Reporting on quality of life in randomized controlled trials: Bibliographic study. BMJ 1998; 317: 191–194.
16. Sullivan M. The new subjective medicine: Taking the patient's point of view on health care and health. Social Sci Med 2002; 56(7): 1595–1604.
17. 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: 347–357.
18. Verrips GH, Stuifbergen MC, den Ouden AL, et al. Measuring health status using the Health Utilities Index: Agreement between raters and between modalities of administration. J Clin Epidemiol 2001: 54(5); 475–481.
19. LeGales C, Costet N, Gentet JC, et al. Cross-cultural adaptation of a health status classification system in children with cancer. First results of the French adaptation of the Health Utilities Index Marks 2 and 3. Int J Cancer Suppl 1999; 12: 112–118.
20. Glaser AW, Davies K, Walker D, Brazier D. Influence of proxy respondents and mode of administration on health status assessment following central nervous system tumours in childhood. Qual Life Res 1997; 6(1): 43–53.
21. Sprangers MAG, Aaronson NK. The role of health care provides and significant others in evaluating the quality of life of patients with chronic disease: A review. J Clin Epidemiol 1992; 45: 743–760.
22. Eiser C, Morse R. Quality of life measures in chronic diseases of childhood. Health Technol Assess 2001; 5(4): 1–157.
23. Patrick DL, Erickson P. Health status and health policy: Quality of Life in Health Care Evaluation and Resource Allocation. New York: Oxford University Press; 1993.
24. Feeny D, Leiper A, Barr RD, et al. The comprehensive assessment of health status in survivors of childhood cancer: Application to high risk acute lymphoblastic leukaemia. Br J Cancer 1993; 67(5): 1047–1052.
25. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions: Health Utilities Index. PharmacoEconomics 1995; 7(6): 503–520.
26. Feeny D, Furlong W, Torrance G. The health utilities index: An update. Qual Life Newslett 1999; 22: 8–9.
27. Barr RD, Pai MKR, Weitzman S, et al. A multi-attribute approach to health status measurement and clinical management illustrated by an application to brain tumors in childhood. Int J Oncol 1994; 4: 639–648.
28. Feeny DH, Furlong W, Barr RD, Torrance GW, Rosenbaum P, Weitzman S. A comprehensive multi-attribute system for classifying the health status of survivors of childhood cancer. J Clin Oncol 1992; 10(6): 923–928.
29. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 System. Med Care 2002; 40(2): 113–128.
30. Kaasa S, Hjermstad MJ, Jordhoy MS, Wisloff F, Loge JH. Compliance in quality of life data: A Norwegian experience. Stat Med 1998; 17(5–7): 623–632.
31. Curran D, Bacchi M, Schmitz SF, Molenberghs G, Sylvester RJ. Identifying the types of missingness in quality of life data from clinical trials. Stat Med 1998; 17: 739–756.
32. Fairclough DL, Peterson HF, Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? Stat Med 1998; 17: 667–677.
33. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York, NY: J. Wiley and Sons, 1987.
34. Furlong WJ, Feeny DH, Torrance GW. McMaster Health Utilities Index: Algorithm for determining health status classification levels, health states, health-related quality of life scores and single attribute level utility scores from HUI 2.3 40-item interviewer-administered health status questionnaires, 1998.
35. Velikova G, Stark D, Selby P. Quality of life instruments in oncology. Eur J Cancer 1999; 35(11): 1571–1580.
36. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: Differences between improvement and worsening. Qual Life Res 2002; 11(3): 207–221.
37. Roila F, Cortesi E. Quality of life as a primary end point in oncology. Ann Oncol 2001; 12 (Suppl 3): S3–6.
38. Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory.Med Care 1999; 37(2): 126–139.
39. Varni JW, Burwinkle TM, Jacobs JR, Gottschalk M, Kaufman F, et al. The PedsQL in type 1 and type 2 diabetes: Reliability and validity of the Pediatric Quality of Life Inventory Generic Core Scales and type 1 Diabetes Module. Diabetes Care. 2003; 26(3): 631–637.
40. Raphael D, Rukholm E, Brown I, Hill-Bailey P, Donato E. The quality of life profile-adolescent version: Background, description and initial validation. J Adol Hlth 1996; 19: 366–375.
41. LeCoq EM, Boeke AJP, Bezemer PD, Bruil J, vanEijk JThM. Clinimetric properties of a parent report on their offspring's quality of life. J Clin Epidemiol 2000; 53: 139–146.
42. Cummins R. Directory of Instruments to Measure Quality of Life and Cognate Areas. Melbourne, Australia: Deakin University, 1997.
43. Saigal S, Feeny D, Furlong W, et al. Comparison of the health related quality of life of extremely low birth weight children and a reference group of children at age eight years. J Pediatr 1994; 125: 418–425.
44. Barr RD, Petrie C, Furlong W, et al. Health related quality of life during post inducing chemotherapy in children with acute lymphoblastic leukemia in remission: An influence of corticosteroid therapy. Int J Oncol 1997; 11: 333–339.
45. Barr RD, Furlong W, Dawson S, et al. An assessment of global health status in survivors of acute lymphoblastic leukemia in childhood. Am J Pediatr Hematol Oncol 1993; 15: 284–290.

46. Gemke RJ, Bonsel GJ. Reliability and validity of a comprehensive health status measure in a heterogeneous population of children admitted to intensive care. J Clin Epidemiol 1996; 49(3): 327–333.
47. Boyle MH, Furlong W, Feeny D, Torrance G, Hatcher J. Reliability of the Health Utilities Index – Mark III used in the 1991 Cycle 6 General Social Survey health Questionnaire. Qual Life Res 1995; 1(4): 249–257.
48. Whitton AC, Rhydderch H, Furlong W, Feeny D, Barr RD. Self-reported comprehensive health status of adult brain tumor patients using the Health Utilities Index. Cancer 1997; 80(2): 258–265.
49. Mulhern RK. Correlation of the Health Utilities Index Mark 2 Cognition Scale and neuropsychological functioning among survivors of childhood medulloblastoma. Int J Cancer 1999; Suppl 12: 91–94.
50. Nixon-Speechley K, Maunsell E, Desmeules M, et al. Mutual concurrent validity of the child health questionnaire and the health utilities index: An exploratory analysis using survivors of childhood cancer. Int J Cancer Suppl 1999; 12: 95–105.
51. Barr RD, Chalmers D, DePauw S, Furlong W, Weitzman S, Feeny D. Health related quality of life in survivors of Wilms' tumor and advanced neuroblastoma: A cross-sectional study. J Clin Oncol 2000; 18(18): 3280–3287.
52. Barr RD, Gonzalez A, Longchong M, et al. Health status and health-related quality of life in survivors of cancer in childhood in Latin American: A MISPHO feasibility study. Int J Oncol 2001; 19(2): 413–421.
53. Wang Q, Furlong W, Feeny D, Torrance G, Barr R. How robust is the Health Utilities Index Mark 2 utility function? Med Decis Making 2002; 22(4): 350–358.
54. LeGales, Buron C, Costet N, Rosman S, Slama PR. Development of a preference-weighted health status classification system in France: The Health Utilities Index 3. Health Care Manage Sci 2002; 5(1): 41–51.
55. Silberfeld M, Rueda S, Krahn M, Naglie G. Content validity for dementia of three generic preference based health related quality of life instruments. Qual Life Res 2002; 11(1): 71–79.
56. Smith KW, Avis NE, Assmann SF. Distinguishing between quality of life and health status in quality of life research: A meta-analysis. Qual Life Res 1999; 8(5): 447–459.
57. Trudel JG, Rivard M, Dobkin PL, Leclerc JM, Robaey P. Psychometric properties of the Health Utilities Index Mark 2 system in paediatric oncology patients. Qual Life Res 1998; 7(5): 421–432.
58. Oostenbrink RA, Moll HA, Essink-Bot ML. The EQ-5D and the Health Utilities Index for permanent sequelae after meningitis. A head-to-toe comparison. J Clin Epidemiol 2002; 55(8): 791–799.
59. Kopec JA, Williams JI, To T, Austin PC. Cross-cultural comparisons of health status in Canada using the Health Utilities Index. Ethnic Health 2001; 6(1): 41–50.
60. Kopec JA, Schultz SE, Goel V, Ivan-Williams J. Can the health utilities index measure change? Med Care 2001; 39(6): 562–574.
61. Dlugos D, Ravina B, Jacobstein C, Chimwaza A, Siderowf A. Comments on: Applicability of the Health Utilities index to a population of childhood survivors of central nervous system tumours in the UK. Glaser AW, Furlong W, Waltzer DA, et al. Eur J Cancer 1999; 35: 256–261; Eur J Cancer 2000; 36(11): 1459–1460.
62. Jenkinson C. Death by questionnaire: Quality of life measurement could seriously damage your health. J Health Serv Res Policy 1999; 4(3): 129–130.
63. Kennedy CR, Leyland K. Comparison of screening instruments for disability and emotional behavioral disorders with a generic measure of health related quality of life in survivors of childhood brain tumors. Int J Cancer 1999; Suppl 12: 106–111.
64. Younossi ZM, Boparai N, McCormick M, Price LL, Guyatt G. Assessment of utilities and health-related quality of life in patients with chronic liver disease. Am J Gastroenterol 2001; 96(2): 579–583.

Address for correspondence: Cheryl L. Cox, St. Jude Children's Research Hospital, 332. Lauderdale St. Memphis, TN 38105-2794, USA
Phone: +1-901-495-4789; Fax: +1-901-495-2866
E-mail: cheryl.cox@stjude.org