Quality of Life Research (2005) 14: 1613–1625
DOI 10.1007/s11136-005-1004-2

Preliminary psychometric evaluation of the Child Health Ratings Inventory (CHRIs) and Disease-Specific Impairment Inventory-Hematopoietic Stem Cell Transplantation (DSII-HSCT) in parents and children

S.K. Parsons[1,2], M.C. Shih [3,4], D.K. Mayer[1,2], S.E. Barlow[5], S.E. Supran[1], S.L. Levy[6], S. Greenfield[7] & S.H. Kaplan[7]
[1]*Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center (E-mail: sparsons@tufts-nemc.org)* [2]*Tufts University School of Medicine;* [3]*Clinical Research Program, Children's Hospital;* [4]*Harvard School of Public Health, Boston MA;* [5]*St. Louis University School of Medicine, St. Louis, MO;* [6]*University of Massachusetts, Worcester, MA;* [7]*University of California, Irvine, CA*

Accepted in revised form 13 January 2005

Abstract

Purpose: To describe the initial results of the Child Health Ratings Inventory (CHRIs), 20-item generic health-related quality of life (HRQL) instrument and the 10-item disease-specific (DS) module, the Disease Specific Impairment Inventory-Hematopoietic Stem Cell Transplantation (DSII-HSCT), for children and adolescents, ages 5–18 years and their parents following HSCT. *Study design:* Using cross sectional design, 122 children with a median age of 11 years (range 5.0–18 years) completed the questionnaire (CHRIs + DSII-HSCT) with research assistance. Seventy-four parents independently completed a parallel version of the questionnaire; health care providers assigned a global clinical severity rating. *Results:* The generic core includes four domains: **physical**, **role**, and **emotional functioning**, and **energy**. The DS module has three domains: **worry, hassles**, and **body image**. The Cronbach's alpha for parents and for older children (8 years and over) exceeded 0.70 for all generic and DS domains. While the range of alpha coefficients was lower for younger children, ages 5–7 year, only the alpha coefficient for one domain (energy) was less than 0.70. The instrument satisfactorily discriminated between clinically important groups: those early in the transplant process (< 6 months) versus those later (> 12 months) and by provider-assigned clinical severity ratings. *Conclusion:* results suggest that the CHRIs generic core and its DSII-HSCT module is a promising measure of HRQL after pediatric HSCT. Although parent and child reports were moderately correlated and revealed complementary results, the unique perspectives of both raters provide a more complete picture of HRQL. Longitudinal application is underway to further characterize the measurement properties of the CHRIs and to determine the instrument's responsiveness and sensitivity to change over time in this vulnerable population.

Key words: Children's self-assessment, Health-related quality of life, Health status, Hematopoietic stem cell transplantation, Parent report, Pediatric oncology

Introduction

Quality of life (QOL) is a subjective, multidimensional concept that encompasses a variety of domains, including physical health, psychological state, levels of functional independence, social relationships, environmental features, spiritual concerns, and personal beliefs [1]. The terms "QOL" and "health-related quality of life" (HRQL) are often used interchangeably, although the latter refers to the impact of health states on overall QOL. The principal distinction lies in the consideration of selected aspects of overall QOL, such as economic, spiritual, political, and cultural

factors, which are not under the purview of the health care system [2]. Hence, they are not considered in many operationalized definitions of HRQL [3, 4].

Measurement of HRQL is important in understanding the impact of illness and treatment, as a screening tool to identify functional impairment, and in evaluating intervention strategies. Pediatric HRQL instrument development, lagging behind efforts for adults, has a variety of distinct methodological issues that require special attention, [5–7] including the impact of physical and psychosocial development on children and adolescents, the validity of self-report and proxy report, and the need for age-appropriate item content [5, 8].

Historically, it was thought that children's developing emotional maturity and cognitive capacity rendered them incapable of providing consistent and accurate information about their level of functioning or state of well-being [9–11]. The design of developmentally appropriate (and age-specific) measures that are accessible to children across a wide age range represents one of the most important innovations in this area and serves as a model for the work reported herein [7, 12–14]. Moreover, there has been growing awareness of the developmental skills required for valid child self-report; these in turn have informed instrument development and evaluation [6, 7, 9–11, 13, 14]. Briefly, these composite skills include the ability to conceptualize health and make the linkage between health and functioning. The rater must understand health-related words and grasp complex concepts. With the evolution of abstract reasoning, the child also develops the ability to understand causality/relatedness, have independent thought, and have an expanded capacity for a range of response rather than polar extremes. The ability to recall a specific reference period also evolves.

HRQL instrument development has relied on two distinct approaches: the utility-based measures such as the Health Utilities Index [15, 16] with its roots in economics theory; and the psychosocially-oriented profile instruments, including the Rand Health Insurance Experiment-Child Form [3], the Child Health and Illness Profile (CHIP) [4, 14, 17, 18], the Child Health Questionnaire (CHQ) [19], the Child Health Ratings Inventory (CHRIs) [20], and the PedsQL [21]. Some of these instruments are designed for a targeted age group (e.g., CHIP adolescent and children versions), while others have age-based modules, (e.g., PedsQL) addressing the distinct developmental issues of children in early and middle school age as compared to later adolescence. Disease specific (DS) instruments have also been developed to address the unique symptoms, disruptions and dysfunction associated with discrete conditions [22–25]. Generic cancer-specific measures, namely the PedsQL Cancer Module [12] and the Minneapolis-Manchester Quality of Life Survey (MM-QOL) [26, 27] recently have become available, each with a specific emphasis, focusing on anxieties and discomfort related to medical encounters and issues of survivorship, respectively.

Children who have undergone hematopoietic stem cell transplant (HSCT) are at particular risk for altered HRQL. This is in part due to the paradox of HSCT: the intense treatment that offers patients and their families hope against fairly certain death from their disease may actually result in death itself [28] and invariably leads to acute and delayed sequelae. The meaning of this 'risking death to live', the 'second opportunity to live' and the related expectations of the post transplant period influence the way in which HRQL is perceived by the rater in this population [28]. HSCT also dramatically disrupts children's development both acutely during the hospitalization and subsequently during the extended period of protective isolation. When patients attempt to resume their age-appropriate activities, some find that their developmental trajectory and future potential have been permanently altered. Moreover, the enormous stress of the underlying disease and its treatment and the fear of sequelae, relapse, and death has the potential to alter relationships within the family.

While traditional sources of information, such as clinical outcomes, healthcare utilization, or reports of school absence [29], provide us with some insight about the issues of recovery following HSCT, a comprehensive evaluation of HRQL in children following HSCT has remained elusive. Recent measurement work by Phipps and co-workers in the pediatric HSCT population has addressed HRQL in the acute peri-transplant periods, particularly the impact of transplant on emotional distress and impaired functioning [30]. However, the ability to assess HRQL in school-aged children and adolescents at different time

points beyond the peri-transplant period has been limited by the lack of available instruments.

To address this need, we set out to develop a DS, child-reported instrument for the pediatric HSCT population for use with a conceptually similar general health status measure, the CHRIs [20], which was also under development at that time. The CHRIs instruments permit comparison between self- and proxy reporting through the use of parallel versions for parents and children (both school-aged and adolescents), utilizing a generic core and a DS module. This paper is the report of the preliminary testing of the CHRIs instruments among pediatric HSCT survivors and their parents.

Methods

Study participation

Pediatric patients, age 5–18 years, receiving post-transplant care at the Jimmy Fund Clinic of the Dana-Farber Cancer Institute (DFCI) following HSCT at Children's Hospital, Boston, and their parents were consecutively recruited for study participation. Criteria for participation included an age-appropriate working knowledge of English at the time of the interview; functionally corrected hearing and vision, and a routine, scheduled visit with their regular physician/nurse practitioner (NP). All eligible patients were identified from the daily clinic roster and, together with their parents, were recruited upon their arrival to the clinic by a member of the research team. Written consent from parents and verbal assent from children were obtained according to the policies of the hospital's institutional review board.

Study design and data collection

Questionnaires for the child and parent, described below, were completed at the end of a routine visit. To establish the link between health status and clinical status, a medical chart review was performed on all participants. The various components of the transplant process (e.g., transplant type, time after transplant, presence or absence of chronic graft versus host disease (CGVHD) [31] serve as markers of clinical severity, informed by the degree of immunosuppression, infection risk, risk of complications, and extent of protective isolation – all of which could have functional impact [32]. The relationship of clinical severity and HSCT is supported by several studies in adult HSCT recipients [33–44].

Health care provider severity rating: Following a scheduled visit, physicians or NP completed a global severity item, ''How would you rate the severity of this child's condition post-HSCT, overall?'' The overall severity item was scored on a 5-point scale from ''minimal'' to ''very severe'' to determine the disease severity rating (DSR).

Development of the CHRIs DSII-HSCT instrument

Development of child self-report: The school aged-version of CHRIs was originally developed for children age 5–12 years with chronic diseases as part of a larger initiative led by Dr Sherrie Kaplan to enhance children's understanding of and participation in their own disease management [20]. This generic instrument consists of 20 items with developmentally appropriate item content, forming four domains: physical, emotional and role functioning, and energy. The adolescent version, developed as part of the HSCT research efforts, contains the same core items. The specific content of HSCT questions for the DS module was based on focus groups, semi-structured interviews and polling of patients, parents, and health care providers as well as a formal review of the literature to derive the most salient aspects of the transplant experience. Ten items were developed within three hypothesized domains: hassles, body image, and worry (i.e., distress/preoccupation) for each age-based version and for parents. The resultant questionnaire was reviewed with elementary school-aged and adolescent transplant recipients, their parents, and their health care providers for face validity, including completeness, relevance, and perceived burden.

Children completed the questionnaire away from their parents to avoid prompting or influence. For younger children, particularly those who were pre-literate, the questionnaire was completed with a research assistant working from a standardized script to avoid bias. For other children, research assistance was available as needed. The entire questionnaire (general and DS modules)

was completed in 10–25 min with the longer times reflecting interviewer assistance for the younger children.

Development of parental report: Parents completed a parallel questionnaire privately, although the research assistant was available, if needed. The parental questionnaire includes three sections: (1) questions identical to those posed to the child, regarding the child's general health and DS functioning; (2) questions on the parent's *own* health and level of functioning; (3) questions on the family's demographic characteristics. As a result, the parents completed the assessment in 20–25 min, inclusive of extensive, unsolicited comments about the transplant process and their thoughts about the questionnaire.

The generic and DS questions were designed to assess children's status in the four weeks before the interview. The voice of the child is the standard used in the questions. All questions began with the stem "How much has not feeling well gotten in the way…" followed by a description of an area of general (CHRIs) or DS functioning (DSII-HSCT). For the school-aged children, the response scales were pictorially represented, and the child was instructed to select "the kid most like you" from among the five response levels. In contrast, the five response options were text-based in the adolescent and parent versions; adolescents had reported finding the pictures too "baby-ish" when probed during the instrument's developmental phase. For frequency items, the 5-point Likert scale ranged from "never" to "a whole lot of the time," whereas for the intensity items, the scale ranged from "not at all" to "a whole lot." For the DS items, the 5-point rating scale ranged from most (1) to least (5). For most of the DS items, the respondent also was given the option of indicating that "this hasn't happened to me" (or my child) or "I don't have to do this." This choice was designed to distinguish between experiencing the problem and not being bothered by it *versus* not experiencing the problem within the reference period. The scores for each general and DS dimension were then transformed to a 0-100-point scale[1] using conventional psychological scaling methods [45]. For the generic items the higher number represents a higher level of reported functioning, while for the DS items, the higher score connoted greater distress, more hassles, or body image issues.

[1] The details of transformation of raw scale scores are outlined in the manual and interpretation guide of the SF-36 Health Survey, 1993, p. 6:18.

Statistical analysis

Sample size

The planned sample ($n = 50$/group) was chosen to yield an 80% power to detect meaningful differences in the mean summary scores for each domain. A medium effect of 7–10 points between groups [46] was chosen and the sample size calculation was based on the observed variance in responses from children with other chronic diseases as well as preliminary data from a sample of healthy children (S.H. Kaplan, personal communication). This sample size was selected to address the primary objective of the study, which was to demonstrate "proof of concept" in the evaluation of HRQL in the HSCT pediatric population.

To characterize the psychometric properties of CHRIs and the DSII-HSCT within this transplant sample, Cronbach's alpha coefficient was computed to evaluate the internal consistency reliability for each of the four domains in the general measure and each of the three domains the DS module with a minimum acceptable criterion of 0.70 [47]. The relationship between individual items to the scales was described by reporting the Pearson correlation for each item to its respective subscale with the item itself removed. Summary scores for each generic and disease-specific scale were defined as the average raw score if at least 50% of the items in the scale were non-missing; if more than 50% of the items were missing, the scale was set to missing. This is equivalent to imputing the missing values(s) by the mean of the available items when at least 50% of the items were not missing. This calculation was done for both raters (children and parents).

To evaluate the discriminant validity of the measure, mean scores in generic and DS scales were compared using known-groups methodology by time post transplant. Specifically, we compared the pattern of scores from each rater for those early in the post-transplant period (< 6 months,

i.e, **EARLY,** 38 children and 23 parents) with those who were more than six months post transplant (i.e., **LATE**, 84 children and 51 parents), by two-sample *t*-test**.** We hypothesized that transplant recipients in the **EARLY** group would report more disease-related hassles, worries, and changes to body image than those in the **LATE** group, based on externally imposed protective isolation and more intense transplant-related management. We also hypothesized that physical functioning and energy functioning would be more severely affected in the **EARLY** group than in the **LATE** group. All tests were two-sided. Of note, we also performed comparisons that adjusted for child age at assessment, as a dichotomous variable (5–7 years vs. 8 years and older) or as a continuous variable. These adjusted analyses yielded the same pattern of scoring and comparable magnitude of *p*-values and thus not presented.

As both a measure of discrimination as well as of convergence, the mean summary scores for each rater (parent and child) were compared with respect to health care provider-assigned clinical severity ratings. The analysis focused on children who were at least 6 months post-transplant. The severity rating's five levels were collapsed into three, based on the distribution of responses: level 1, minimal (49.6%); level 2, mild-moderate (41.2%); and level 3, severe-very severe (9.2%). Results were adjusted for patient age and time post transplant. Pairwise comparisons were performed (e.g., minimal vs. severe); Tukey adjusted *p*-values are reported [48].[2]

For comparisons between raters within the treatment dyad, the mean differences (child score minus parent score) and the standard errors were calculated for each generic and DS domain. Multiple linear regressions were performed to examine the relationship between the mean difference in domain scores and the child's age at the assessment, time since transplant, and transplant type. Age and time since transplant were considered both dichotomously and continuously, with the final selection based on goodness of fit statistics and Akaike Information Criterion. We also examined potential interactions between raters and other covariates; that is, whether parent–child differences in domain scores varied with respect to other covariates. Furthermore, as a measure of inter-rater reliability, intra-class correlations (ICC) were calculated using a mixed effects model with all available data [49]. Pearson's correlation was also calculated as a measure of inter-rater validity [49]. Of note, neither the ICC nor the Pearson's correlation was adjusted for covariates such as child's age at the assessment or time since transplant, based on the premise that parent's and child's reports may co-vary *because of* these covariates.

[2] The reader is referred to an excellent review by Gerald E. Dallal, PhD, entitled Multiple Comparison Procedures, available at: www.tufts.edu/~gdallal/mc.htm, accessed 11/5/2004.

Results

Sample description

A convenience sample of 122 children was enrolled, 48 without corresponding parent data and 74 in treatment dyads. Eligible participants were approached for recruitment, based on consecutive presentation to the clinic for routine post-transplant care. Among those, two parents and two children in four dyads refused to participate, resulting in a 1.6% refusal rate for the children (2/122) and 2.7% for parents (2/74). The two children who refused to participate reported that they did not want to think about their transplant; both were >12 months post-transplant. Both parents did not want their children to participate. In addition, four children were unable to complete the assessment, despite research assistance and attempts to modify attentional and/or spatial confounders. After the initial pilot testing of the measure in school aged children ($n = 48$), the DS module for the parent questionnaire was augmented to include all 10 DS items to ensure complete parallel content and uniformity of scoring with the child version. Hence, the parent results reported herein reflect 74 respondents who completed the augmented parent questionnaire. Since no modifications of item content were made in the child measure after the first pilot, the results for the entire 122 patient sample are included.

Demographics of the study sample are summarized in Table 1. Overall, the sample was predominantly Caucasian (94%) and male (54%), reflecting the overall HSCT population at our

Table 1. Sample demographics (74 parents and 122 children)[a]

Child gender	54% Male
Child race	94% Caucasian
Child age	
Age at BMT median (range)	8 (1–17)
Age at Eval median (range)	11 (5–18)
Transplant indication	77% hematologic malignancies
Transplant type	58 (48%) autologous 64 (52%) allogeneic
Days post BMT median (range)	519 (17—3079)
% with CGVHD*	29% (12/41)
Parental education	
Did not completed high school	6%
Completed high school	16%
Some–completed college	63%
Some–completed graduate school	15%
Parental marital status	
Married/living with partner	87%
Divorced/separated/widowed	12%
Never married	2%
Median family income	$40–49,999
Method of payment	
Commercials insurance	36%
HMO/prepaid plan	23%
Other/combo**	25%
Medicaid	14%
Self pay	2%

[a] Numbers for gender, parental education, and marital status are not complete due to missing data.
* Allo transplants > 100 days.
** Reflects more than one type of private insurance coverage.

center. Gender differences reflect the epidemiology of the underlying conditions. The median age at evaluation was 11 years. Although the median time post-transplant was one and a half years, (519 days), the range extended from 17 days to 8.4-years (3079 days). The distribution of the study sample by diagnosis and transplant type mirrored that of the center's overall HSCT population with a predominance of transplants for hematological malignancies (77%, study; 69% center overall) and near-equivalence in the number of allogeneic (allo) and autologous (auto) transplants (52% allo; 48% auto). Among the 41 allo transplant recipients greater than 100 days post-transplant and thus evaluable for CGVHD, 12 (29%) had CGVHD, reflecting the prevalence of this complication in our center. The families of participants were well educated (78% with more than a high school education) with a median family income of $40,000–49,999. While many were privately insured, 14% of the study population had Medicaid coverage, similar to the program as a whole (16%).

Measurement properties

1. *Item scores*

Mean raw scores (standard deviation) for each item, presented in Tables 2 and 3, revealed similar variability between parent and child reports. All generic and DS items demonstrated sufficient variability across response choices. Ceiling and floor effects were not observed for either rater; the full five-point scale was used in the majority of items. Among the generic items, none of the parent raters endorsed that the child was doing "pretty bad" lately or had "a whole lot" of pain. None of the child raters selected that they were "sad." In the DS items, none of the parents reported that the child was bothered "a whole lot" by telling friends about the transplant.

2. *Internal Consistency Reliability (ICR)*

The results of the internal consistency reliability, summarized in Table 4 reveal that alpha coefficients for the generic domains range from 0.85 to 0.95 for the parent raters and 0.69–0.83 for children for the generic items. As expected, younger children, age 5–7 years, had lower reliability for all scales. Among the older children (8 years and older), alpha coefficients ranged from 0.72 to 0.86. Similar patterns were observed for the DS module. Of note, since the body image domain contains only two items, an alpha coefficient was not calculated.

3. *Validity*

The unadjusted comparison of mean summary scores by time post transplant (early vs. late) is summarized in Table 5 for each domain by rater. For all domains, generic and DS, early scores from the parents are significantly worse than later scores ($p < 0.001$). The magnitude of the differences in mean summary scores for the two time periods is striking, ranging from 20 to 50 points (1–2 SD). This pattern is also seen in the child reports for all DS domains and for energy. For the other generic domains (i.e., physical, role, emotional), early scores are lower than later scores, but those differences do not reach statistical significance.

The comparison of domain scores with the health care provider's disease severity rating is

Table 2. Generic item mean raw scores by rater (1 = worst functioning to 5 = best functioning)

Item	Parent (n = 74)			Child (n = 122)		
	Mean (SD)	Floor%	Ceiling%[a]	Mean (SD)	Floor%	Ceiling%[a]
Play hard	3.44 (1.59)	20.8	38.9	3.44 (1.30)	10.1	25.2
Play ball	3.64 (1.57)	15.9	49.3	3.73 (1.30)	9.2	36.7
Swing/walk	3.96 (1.35)	8.2	53.4	4.03 (1.13)	5.0	44.5
Climb	3.74 (1.54)	14.5	52.2	3.46 (1.40)	13.3	31.7
Class activities	3.79 (1.65)	20.6	58.7	3.81 (1.41)	13.3	46.0
Housework	4.01 (1.37)	10.8	55.4	3.76 (1.36)	11.1	43.6
Schoolwork	3.96 (1.32)	7.1	52.9	3.85 (1.29)	8.0	45.1
Enjoy family	4.05 (1.28)	5.5	57.5	4.14 (1.24)	6.8	56.4
Pay attention	4.21 (1.20)	4.0	63.0	3.82 (1.23)	6.8	39.3
Concentrate	4.28 (1.31)	8.1	71.6	4.04 (1.28)	8.5	50.9
Talk friends	3.96 (1.46)	11.1	59.7	4.13 (1.23)	6.7	57.1
Miss school	2.70 (2.29)	1.4	38.6	2.60 (2.12)	7.8	27.6
Energy after school	2.54 (2.01)	2.9	15.9	3.57 (1.44)	7.1	27.4
Energy play	3.60 (1.32)	13.9	27.8	4.14 (1.00)	1.7	46.2
Rest	3.64 (1.06)	1.4	26.0	3.89 (1.09)	5.1	35.6
Nervous/worry	3.53 (1.01)	2.8	15.3	4.07 (1.05)	4.2	41.2
Happy/sad	3.46 (1.05)	1.4	16.9	4.06 (0.96)	0.0	40.3
Doing overall	4.09 (0.93)	0.0	40.0	4.40 (0.88)	0.9	61.5
Fun	3.74 (1.06)	1.4	29.7	4.09 (1.09)	3.4	48.7
Pain	4.14 (0.93)	0.0	43.8	4.24 (1.01)	3.4	52.9

[a] Floor: worst function; ceiling: best functioning.

summarized in Table 6. In this comparison, the best model fit included age as a continuous variable and time post-transplant, based on quartiles. It revealed that parents reported significantly lower scores in emotional functioning for those in the most severe category as compared to the least (minimal) severe group. Similarly, parental scores for energy significantly varied for those in the most severe group versus minimal, but also in the mild-moderate versus most severe. This pattern was also seen for body image. Mirroring parent responses, children reported significantly different scores in comparisons between the least and most severe groups for physical functioning and all three of the disease specific domains.

4. *Inter-rater comparisons*

Table 7 summarizes the results of a multiple linear regression examining mean difference (child score minus parent score) for each of the generic and

Table 3. Disease specific item mean raw scores by rater (1 = least impaired to 5 = most impaired)

Item	Parent (n = 74)			Child (n = 122)		
	Mean (SD)	Floor %	Ceiling %[a]	Mean (SD)	Floor %	Ceiling %[a]
Worry hosp	2.00 (1.30)	52.7	6.8	1.84 (1.14)	53.9	4.3
Worry infection	1.93 (1.21)	51.4	5.6	1.99 (1.31)	50.9	9.3
Worry disease	2.01 (1.32)	52.8	6.9	1.95 (1.32)	53.4	10.2
Mask	1.38 (1.52)	43.5	8.7	1.65 (1.55)	37.1	6.9
Diet	1.41 (1.44)	36.2	5.8	1.96 (1.63)	33.6	14.0
Meds	1.60 (1.41)	34.3	5.7	1.76 (1.35)	43.6	5.5
Tell friend	1.37 (0.67)	72.1	0.0	1.73 (1.47)	51.7	8.6
Think BMT	2.65 (0.99)	13.5	2.7	1.99 (0.99)	35.6	3.4
Cheek	0.97 (1.18)	40.0	1.4	1.39 (1.40)	40.4	4.4
Hair	1.27 (1.30)	44.3	4.3	1.46 (1.45)	45.2	7.0

[a] Floor: least impaired; ceiling: most impaired.

Table 4. Internal consistency reliability (Cronbach's alpha)

	Number of item items	Parent (n = 74)	Child		
			Child (all) (n = 122)	5–7 years (n = 32)	8–18 years (n = 90)
Generic					
Physical	6	0.95	0.82	0.79	0.83
Role	5	0.91	0.83	0.74	0.86
Emotional	5	0.85	0.73	0.72	0.74
Energy	4	0.89	0.69	0.59	0.72
Disease specific					
Worry	3	0.91	0.77	0.69	0.80
Hassle	4	0.84	0.74	0.70	0.77
Body image	2	–	–	–	–

disease specific domains. The parent scores varied significantly by time post-transplant in all domains ($p < 0.0001$). For example, physical functioning scores were 40 points higher (SE 6.9) for the group beyond 6 months versus less than 6 months post-transplant. Scores for Hassles were 29 points lower (SE 4.3), connoting fewer hassles, in the later time period. A similar pattern was noted for all other domains (see Table 7). The impact of child age on the mean difference varied by domain. For emotional functioning and worry, emotional distress significantly increased with child age. This was reflected in lower emotional functioning scores (coefficient of −1.12; SE 0.37) and higher worry scores (coefficient 2.15, SE 0.78) when age was considered continuously. Child age was not significantly associated with other domain scores. When interactions between rater and other covariates (time post-transplant, transplant type, and child age at evaluation) were considered, several important patterns emerged. First, for all of the generic and disease-specific domains, the interaction between rater and time post-transplant was significant. As an example, for physical functioning, in the early period (< 6 months), children's scores were 19.9 points higher than parent scores ($p = 0.003$), whereas in the later period, parents' scores were 12.2 points higher than the children's scores ($p = 0.005$). Second, for energy, the rater by transplant type interaction was also significant. Specifically, parents of allo transplant recipients reported 8.2 points worse energy than parents of auto recipients ($p = 0.05$), whereas children's scores did not significantly vary by transplant type ($p = 0.94$). Third, a significant interaction between rater and child age at assessment was observed for the disease-specific domain of worry. Specifically, the association between age at assessment and

Table 5. Parent and child: early (< 6 months) and late (≥ 6 months) mean domain scores

	Parent				Child			
	Total (n = 74) Mean (SD)	Early (n = 23) Mean (SD)	Late (n = 51) Mean (SD)	*p*-value	Total (n = 122) Mean (SD)	Early (n = 38) Mean (SD)	Late (n = 84) Mean (SD)	*p*-value
Generic								
Physical	69.2 (33.5)	41.7 (33.6)	81.3 (25.5)	< 0.001	67.2 (24.2)	62.4 (20.1)	69.4 (25.7)	0.14
Role	77.6 (28.1)	60.1 (26.6)	85.7 (25.1)	< 0.001	74.8 (24.6)	75.9 (21.2)	74.2 (26.2)	0.74
Emotional	69.6 (19.3)	54.9 (18.1)	76.2 (16.0)	< 0.001	79.2 (17.5)	75.6 (18.8)	80.9 (16.7)	0.13
Energy	53.3 (29.8)	16.6 (9.8)	69.2 (19.7)	< 0.001	60.8 (21.8)	39.7 (11.3)	70.7 (18.2)	< 0.001
Disease specific								
Worry	24.9 (29.6)	48.7 (33.2)	14.0 (20.0)	< 0.001		23.2 (26.0)	30.3 (27.8)	19.8 (24.6)
Hassle	33.8 (19.8)	55.0 (16.2)	23.5 (11.2)	< 0.001	36.1 (20.4)	49.7 (19.9)	29.8 (16.8)	< 0.001
Body image	22.4 (20.8)	41.7 (21.0)	13.0 (12.7)	< 0.001	28.6 (25.1)	37.8 (23.0)	24.3 (25.0)	0.01

Table 6. Health care provider disease severity rating and mean summary scores by domain[a]

	Parent			Child		
	Minimal (n = 26) Mean (SD)	Mild–Moderate (n = 19) Mean (SD)	Severe–very severe (n = 4) Mean (SD)	Minimal (n = 48) Mean (SD)	Mild–moderate (n = 24) Mean (SD)	Severe–very severe (n = 10) Mean (SD)
Generic						
Physical	89.8 (19.4)[b]	73.1 (27.6)	61.5 (37.9)	75.8 (22.4)	63.1 (28.3)	52.9 (24.5)[d]
Role	91.8 (20.9)	80.0 (29.4)	71.3 (28.1)	76.8 (24.5)	70.0 (30.0)	73.9 (23.0)
Emotional	79.6 (14.2)	75.5 (15.2)	58.8 (24.6)[d]	83.7 (15.3)	80.0 (14.6)	72.1 (23.5)
Energy	75.2 (15.7)	67.0 (19.7)[c]	41.3 (26.6)[d]	75.1 (16.1)	66.7 (20.7)	57.2 (16.4)
Disease specific						
Worry	10.0 (18.0)	12.7 (20.9)	31.3 (15.8)	15.4 (19.7)	23.9 (28.7)	36.1 (30.9)[d]
Hassle	21.0 (9.5)	22.6 (10.9)	34.0 (13.3)	27.6 (13.9)	29.0 (18.8)	43.0 (22.3)[d]
Body image	11.2 (11.3)	9.4 (10.0)[c]	30.0 (14.1)[d]	20.7 (22.3)	22.6 (24.3)[b]	44.4 (33.6)[d]

[a] Tukey adjusted p-values ≤ 0.05.
[b] Between minimal/mild–mod.
[c] Between mild–mod/severe.
[d] Between minimum/severe.

worry was observed in parent's scores ($p = 0.008$) but not in child's scores ($p = 0.94$).

Unadjusted intra-class correlations of the raters' scores varied by domain. For the generic domain, the strongest correlation was for energy (0.68), weak to moderate for role, physical and emotional (0.12, 0.26, and 0.33, respectively). In the disease specific domains, the ICCs were in the moderate

Table 7. Child-parent mean differences in HRQL scoring[a]

	Physical coefficient (SD)	Role coefficient (SD)	Emotional coefficient (SD)	Energy coefficient (SD)	Hassles coefficient (SD)	Worry coefficient (SD)	Body image coefficient (SD)
Time post-transplant[b]	40.0 (6.9)***	26.4 (6.6)***	20.4 (4.3)***	47.6 (4.4)***	−18.1 (3.5)***	−31.8 (6.5)***	−22.9 (5.6)***
Age at assessment	−0.4 (0.6)	−0.1 (0.5)	−1.1 (0.4)**	0.3 (0.4)	−0.4 (0.4)	2.2 (0.8)**	−0.8 (0.5)
Transplant type[c]	2.2 (4.5)	3.5 (4.3)	−1.1 (3.0)	−8.2 (4.1)*	2.5 (2.9)	−1.5 (4.6)	5.4 (4.1)
Rater[d]	19.9 (6.4)**	15.4 (6.4)*	19.4 (3.9)***	16.1 (4.7)**	−5.4 (3.9)	11.3 (11.4)	−3.8 (4.4)
Time post-transplant by Rater	−32.1 (7.7)***	−27.0 (7.7)***	−16.4 (4.7)***	−17.0 (4.3)***	−10.9 (4.7)*	19.8 (6.6)**	11.5 (5.3)*
Transplant type by Rater	–	–	–	8.5 (4.0)*	–	–	–
Age of assessment by Rater	–	–	–	–	–	−2.1 (0.8)*	–
Child–parent (early)	19.9 (6.4)**	15.4 (6.4)*	19.4 (3.9)***		−5.4 (3.9)	11.3 (11.4)	−3.8 (4.4)
Child–parent (late)	−12.2 (4.2)**	−11.6 (4.3)**	3.1 (2.6)		5.5 (2.7)*	31.1 (10.7)**	7.7 (3.0)*
Child–parent (early, auto)				16.1 (4.7)**			
Child–parent (early, allo)				24.6 (3.5)***			
Child–parent (late, auto)				−0.9 (2.8)			
Child–parent (late, allo)				7.6 (3.1)*			
Allo–auto (parent)				−8.2 (4.1)*			
Allo–auto (child)				0.3 (3.5)			
Age at assessment (parent)						2.2 (0.8)**	
Age at assessment (child)						0.0 (0.7)	

[a] p-value: * < 0.05; ** < 0.01; *** < 0.001.
[b] Early (< 6 months) vs. late (≥ 6 months).
[c] Reference group = auto (vs. allo) transplant.
[d] Reference group = parent (vs. child).

range (0.43, worry; 0.52, hassles, and 0.60, body image). Pearson's correlations were also performed, yielding similar, but not identical results to the ICC. (Data not shown.)

Discussion

This study outlines the preliminary work in the evaluation of HRQL in children undergoing HSCT, utilizing the CHRIs instruments: a generic core and a DS module with child and parent report versions. The potential advantages of the CHRIs instruments in this population include its brevity, the conceptual links between general and DS modules, and the ability to assess self- and proxy reporting through parallel content questionnaires. An added advantage of the CHRIs instruments for this population is a shared theoretical construct and working definition of QOL with several adult-based instruments (e.g., EORTC QLQ-30 [50], FACT-BMT) [33]. These features make it attractive in complex clinical settings and create the possibility for longitudinal exploration across developmentally distinct age groups.

Both self-report versions of the instrument for school aged children 5–12 years and for adolescents, 13–18 years, contain developmentally appropriate content, presented to enhance the accessibility and acceptability of the measure across this age continuum. Specifically, since the school-aged population represents the majority of the HSCT population and is a developmentally distinct group from its adolescent counterpart, we designed a pictorially based response scale for all general and DS items. This type of scale, which allows the child to match their level of functioning with a picture in the response scale, helps circumvent issues such as literacy and concrete thinking. Other investigators have also successfully used non-verbal cueing with younger children. (Wolters, personal communication, 2004) [12, 14]. As has been previously demonstrated, younger children (5–7 years) had lower internal consistency reliability than older children ($\geqslant$8 years old) [12, 14]. However, with the exception of a single domain, the Cronbach's coefficient exceeded 0.70 for all domains and for all raters.

Our experience mirrors the findings from other pediatric populations. Specifically, when children are provided with the appropriate tool that is relevant and developmentally accessible to them they are capable of providing valid and reliable information [4, 12, 14, 17, 19].

One of the goals of the study was to evaluate the feasibility of assessing HRQL among medically complex children and their parents, receiving care in a busy outpatient clinic. Hence, we were particularly encouraged by the apparent acceptance of the measures as demonstrated by a low refusal rate and high completion rates by both members of the dyad.

One of the most striking results of our study was the magnitude of differences in mean summary scores by time period (early vs. late) and by clinical severity, a difference greater for parent raters than for child raters. Moreover, in addition to magnitude of change, the raters appear to change position as a functioning of time post-transplant. Specifically, children report better functioning than parents in the early post transplant period, whereas later on, parents report better functioning. This pattern is also seen in the disease specific domains. The variability of child response as a function of age most strikingly emerges for the emotional domains (emotional functioning as a generic domain and worry as a disease specific domain), demonstrating a linear relationship between increased distress and child age. Taken together, these results indicate that HRQL differs as a function of time and changing clinical status, mirroring the results from adult transplant recipients. [51, 52]. In addition, these results suggest that the effects of time and clinical status on HRQL have different implications for each rater.

While these initial findings provide compelling information about pediatric HSCT survivors' HRQL, there are several important limitations to this study, principally related to the cross-sectional design, size, and source of the sample. As noted, one of the goals of the study was to demonstrate "proof of concept" that we could evaluate HRQL in this population, using child self report in addition to parent report. We wanted to demonstrate the additional contribution of transplant-specific items as a supplement to items related to generic functioning.

To render the characterization of HRQL in this population more universally applicable, a larger,

more ethnically and socio-economically diverse population is needed. Future research will employ longitudinal analysis to explore potential differences in children's health status as a function of age/developmental stage as well as the differences in raters' perception of HRQL over time. Moreover, we will evaluate the robustness of both the generic core and DS module through direct comparisons with other validated measures. This information will generate a better understanding of the performance of the CHRIs instruments in evaluating HRQL over time in the pediatric HSCT population. With greater confidence we will then be able to identify risk factors leading to impaired functioning for particularly vulnerable patients and families. This knowledge will allow development of appropriate prevention/intervention programs to eliminate or mitigate the long-term effects of HSCT.

Acknowledgements

We gratefully acknowledge the contributions of Allison Moriarty and Lisa Brennan for day-to-day project management and data collection and Amina Khan and Lisa Sullivan for preliminary analytic support. We thank the staff of the Jimmy Fund Clinic at the Dana-Farber Cancer Institute and the patients and families for their participation in this study. Sponsored in part by the Agency for Health Care Policy and Research (HHS R01 HS6897), a Career Development Award from the American Society of Clinical Oncology (SKP), and the American Cancer Society Research Scholar Grant RSGPB 02-186-01-PBP.

References

1. World Health Organization. World Health Organization Constitution. 1947. Geneva, Switzerland: World Health Organization.
2. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA 1995; 273(1): 59–65.
3. Eisen M, Donald C, Ware JE, Brook R. Conceptualization and Measurement of Health for Children in the Health Insurance Study. Santa Monica, CA: Rand Publication, 1990.
4. Starfield B, Riley AW, Green BF, Ensminger ME, Ryan SA, Kelleher K, et al. The adolescent child health and illness profile. A population-based measure of health. Med Care 1995; 33(5): 553–566.
5. Parsons SK, Brown AP. Evaluation of quality of life of childhood cancer survivors: A methodological conundrum. Med Pediatr Oncol 1998; (Suppl 1): 46–53.
6. Eiser C, Morse R. The measurement of quality of life in children: past and future perspectives. J Dev Behav Pediatr 2001; 22(4): 248–256.
7. Rebok G, Riley A, Forrest C, Starfield B, Green B, Robertson J, et al. Elementary school-aged children's reports of their health: a cognitive interviewing study. Qual Life Res 2001; 10(1): 59–70.
8. Parsons SK, Barlow SE, Levy SL, Supran SE, Kaplan SH. Health-related quality of life in pediatric bone marrow transplant survivors: According to whom? Int J Cancer 1999; 12 (Suppl): 46–51.
9. Perrin EC, Gerrity PS. There's a demon in your belly: Children's understanding of illness. Pediatrics 1981; 67(6): 841–849.
10. Perrin EC, Sayer AG, Willett JB. Sticks and stones may break my bones ... reasoning about illness causality and body functioning in children who have a chronic illness. Pediatrics 1991; 88(3): 608–619.
11. Harbeck C, Peterson L. Elephants dancing in my head: A developmental approach to children's concepts of specific pains. Child Dev 1992; 63(1): 138–149.
12. Varni JW, Burwinkle TM, Katz ER, Meeske K, Dickinson P. The PedsQL in pediatric cancer: Reliability and validity of the pediatric quality of life inventory generic core scales, multidimensional fatigue scale, and cancer module. Cancer 2002; 94(7): 2090–2106.
13. Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. Health Technol Assess 2001; 5(4): 1–157.
14. Riley AW, Forrest CB, Rebok GW, Starfield B, Green BF, Robertson JA, et al. The Child Report Form of the CHIP-Child Edition: Reliability and validity. Med Care 2004; 42(3): 221–231.
15. Feeny DH, Furlong W, Barr RD, Torrance GW, Rosenbaum P, Weitzman S. A comprehensive multiattribute system for classifying the health status of survivors of childhood cancer. J Clin Oncol 1992; 10(6): 923–928.
16. Feeny DH, Torrance GW, Furlong WJ. Health utilities index. In: Spilker BF (ed), Quality of Life and Pharmacoeconomics in Clinical Trials, Philadelphia, PA: Lippincott-Raven, 1996: 239–252.
17. Starfield B, Bergner M, Ensminger M, Riley A, Ryan S, Green B, et al. Adolescent health status measurement: Development of the Child Health and Illness Profile. Pediatrics 1993; 91(2): 430–435.
18. Riley AW, Forrest CB, Starfield B, Rebok GW, Robertson JA, Green BF. The parent report form of the CHIP-Child Edition: Reliability and validity. Med Care 2004; 42(3): 210–220.
19. Landgraf JM, Albetz L. Measuring health outcomes in pediatric populations: Issues in psychometrics and application. 2nd ed. Philadelphia, PA: Lippincott-Raven, 1996.
20. Kaplan SH, Barlow S, Spetter D, Sullivan L, Khan A, Grand R. Assessing functional status and health-related

quality of life among school-aged children: Reliability and validity of a new self-reported measure. Qual Life Res 1995; 4: 444.
21. 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.
22. Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in children with asthma. Qual Life Res 1996; 5(1): 35–46.
23. Guyatt GH, Juniper EF, Griffith LE, Feeny DH, Ferrie PJ. Children and adult perceptions of childhood asthma. Pediatrics 1997; 99(2): 165–168.
24. Ingersoll G, Marrero D. A modified quality of life measurement for youths: Psychometric properties. Diabetes Educ 1999; 17(2): 114–118.
25. Feldman BM, Grundland B, McCullough L, Wright V. Distinction of quality of life, health-related quality of life, and health status in children referred for rheumatologic care. J Rheumatol 2000; 27(1): 226–233.
26. Bhatia S, Jenney ME, Bogue MK, Rockwood TH, Feusner JH, Friedman DL, et al. The Minneapolis-Manchester quality of life instrument: Reliability and validity of the adolescent form. J Clin Oncol 2002; 20(24): 4692–4698.
27. Bhatia S, Jenney ME, Wu E, Bogue MK, Rockwood TH, Feusner JH, et al. The Minneapolis-Manchester quality of life instrument: Reliability and validity of the youth form. J Pediatr 2004; 145(1): 39–46.
28. Eilers J, King C. Issues related to bone marrow transplantation. In: King C, Hinds C (eds), Quality of Life from Nursing and Patient Perspectives, Boston, MA: Jones and Bartlett Publishers, 1998: 204–230.
29. Varni JW, Rode CA, Seid M, Katz ER, Friedman-Bender A, Quiggins DJL. The pediatric cancer quality of life inventory-32 (PCQL-32). II. Feasibility and range of measurement. J Behav Med 1999; 22(4): 397–406.
30. Phipps S, Dunavant M, Jayawardene D, Srivastiva DK. Assessment of health-related quality of life in acute inpatient settings: Use of the BASES instrument in children undergoing bone marrow transplantation. Int J Cancer 1999; 12 (Suppl): 18–24.
31. Shulman H, Sullivan K, Weiden P, McDonald G, Striker G, Sale G, et al. Chronic graft-versus-host disease syndrome in man. A long-term clinicopathologic study of 20 Seattle patients. Am J Med 1980; 69: 204–217.
32. Barlow S, Kaplan S, Khan A, Sullivan L, Danford D, Cuttler L, et al. Clinical severity scale development in pediatric chronic disease. Pediatr Res 1996; 39: 130A.
33. McQuellon RP, Russell GB, Cella DF, Craven BL, Brady M, Bonomi A, et al. Quality of life measurement in bone marrow transplantation: Development of the functional assessment of cancer therapy-bone marrow transplant (FACT-BMT) scale. Bone Marrow Transplant 1997; 19(4): 357–368.
34. Bush NE, Haberman M, Donaldson G, Sullivan KM. Quality of life of 125 adults surviving 6–18 years after bone marrow transplantation. Soc Sci Med 1995; 40(4): 479–490.
35. Andrykowski MA, Greiner CB, Altmaier EM, Burish TG, Antin JH, Gingrich R, et al. Quality of life following bone marrow transplantation: findings from a multicentre study. Br J Cancer 1995; 71(6): 1322–1329.
36. Syrjala KL, Chapko MK, Vitaliano PP, Cummings C, Sullivan KM. Recovery after allogeneic marrow transplantation: Prospective study of predictors of long-term physical and psychosocial functioning. Bone Marrow Transplant 1993; 11(4): 319–327.
37. Wingard JR. Functional ability and quality of life of patients after allogeneic bone marrow transplantation. Bone Marrow Transplant 1994; 14 (Suppl 4): S29–S33.
38. Ferrell B, Grant M, Schmidt GM, Rhiner M, Whitehead C, Fonbuena P, et al. The meaning of quality of life for bone marrow transplant survivors. Part 1. The impact of bone marrow transplant on quality of life. Cancer Nurs 1992; 15(3): 153–160.
39. Wolcott DL, Wellisch DK, Fawzy FI, Landsverk J. Adaptation of adult bone marrow transplant recipient long-term survivors. Transplantation 1986; 41(4): 478–484.
40. Altmaier EM, Gingrich RD, Fyfe MA. Two-year adjustment of bone marrow transplant survivors. Bone Marrow Transplant 1991; 7(4): 311–316.
41. Grant M, Ferrell B, Schmidt GM, Fonbuena P, Niland JC, Forman SJ. Measurements of quality of life in bone marrow transplantation survivors. Qual Life Res 1992; 1: 375–384.
42. Schmidt GM, Niland JC, Forman SJ, Fonbuena PP, Dagis AC, Grant MM, et al. Extended follow-up in 212 long-term allogeneic bone marrow transplant survivors. Issues of quality of life. Transplantation 1993; 55(3): 551–557.
43. Haberman M, Bush N, Young K, Sullivan KM. Quality of life of adult long-term survivors of bone marrow transplantation: A qualitative analysis of narrative data. Oncol Nurs Forum 1993; 20(10): 1545–1553.
44. McQuellon RP, Russell GB, Rambo TD, Craven BL, Radford J, Perry JJ, et al. Quality of life and psychological distress of bone marrow transplant recipients: The 'time trajectory' to recovery over the first year. Bone Marrow Transplant 1998; 21: 477–486.
45. DeVellis RF. Scale development: Theory and applications. 2nd ed. Thousand Oaks, CA: Sage, 2003.
46. Cohen M. A power primer. Psychol Bull 1992; 112: 155–159.
47. Cronbach L. Coefficient alpha and the internal structure of test. Psychometrika 1951; 16: 297–334.
48. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. Science 1977; 198(4318): 679–684.
49. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33(1): 159–174.
50. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993; 85(5): 365–376.
51. Bush NE, Donaldson GW, Haberman MH, Dacanay R, Sullivan KM. Conditional and unconditional estimation of multidimensional quality of life after hematopoietic stem

cell transplantation: A longitudinal follow-up of 415 patients. Biol Blood Marrow Transplant 2000; 6(5A): 576–591.
52. Chiodi S, Spinelli S, Ravera G, Petti AR, Van Lint MT, Lamparelli T, et al. Quality of life in 244 recipients of allogeneic bone marrow transplantation. Br J Haematol 2000; 110(3): 614–619.

Address for correspondence: Susan K. Parsons, Director, Center on Child and Family Outcomes, Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center, 750 Washington St., #345, Boston, MA 02111, USA
Phone: +1-617-636-1450; Fax +1-617-636-6280
E-mail: sparsons@tufts-nemc.org