

Comparing longitudinal assessments of quality of life by patient and parent in newly diagnosed children with cancer: the value of both raters' perspectives

Susan K. Parsons · Diane L. Fairclough ·
Jim Wang · Pamela S. Hinds

Accepted: 23 July 2011 / Published online: 6 August 2011
© Springer Science+Business Media B.V. 2011

Abstract

Purpose Health-related quality of life (HRQoL) information from children facing rare and/or life-threatening disease serves important clinical functions. Longitudinal HRQoL ratings from 222 child–parent dyads collected at four time points during the first 16 weeks of cancer treatment are presented. Patient and parent HRQoL reports at the domain level, based on the Pediatric Quality of Life Inventory™ 4.0 Generic Core Scales, were compared over time, and variation in child/parent agreement by age, treatment intensity, and time on treatment was explored.

Patients and methods Analyses included consideration of missingness, differences between child and parent group mean domain scores averaged over assessments, agreement between individual child and parent, compared to group averages, and within-subject changes between assessments.

Results Children consistently reported higher functioning than their parents with differences varying by child age and HRQoL domain and diminishing over time. No differences

were found by intensity of treatment. The between-subject correlation ranged from 0.61 (social functioning) to 0.86 (physical functioning) across time. Agreement within groups, defined by age, treatment intensity, and time were generally similar.

Conclusions Results indicate moderate-to-good child/parent agreement with variability by domain of HRQoL. Findings underscore the complexity of self- and proxy-based report and support the use of information from both raters.

Keywords Pediatric oncology · Patient- and parent-reported health-related quality of life · Patient and parent agreement · Instrumentation · Proxy ratings

Introduction

Over the past 15 years, several validated health-related quality of life (HRQoL) instruments have been developed for child self-report. The availability of companion measures with parallel content for proxy raters, especially parents, has allowed researchers to explore the agreement between raters. Research to date has demonstrated considerable variability in inter-rater concordance, based on child age, gender, diagnosis, duration of illness, treatment status, and by HRQoL domain(s) [1, 3, 7, 9, 10, 14, 27]. Much of this research has been cross-sectional in design, leaving largely unanswered the manner in which patient–parent agreement changes over time and why.

The prevailing wisdom about the role of self-report versus proxy report has shifted dramatically over the past two decades. Prior to the recent proliferation of new measures and validation studies that have accompanied their release, children, especially before adolescence, were

S. K. Parsons (✉)

The Health Institute, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, 800 Washington St., #345, Boston, MA 02111, USA
e-mail: sparsons@tuftsmedicalcenter.org

D. L. Fairclough

Department of Biostatistics and Informatics, Colorado School of Public Health and Colorado Health Outcomes Program School of Medicine, University of Colorado Denver, Denver, CO, USA

J. Wang

J&J Pharmaceutical Research and Development, LLC, Raritan, NJ, USA

P. S. Hinds

Children's National Medical Center, George Washington University, Washington, DC, USA

thought to lack the cognitive/developmental skills to provide valid and reliable self-report. Several emerging instruments have directly addressed these developmental limitations through altered questionnaire design, pictorial response scales, and item wording. Field-testing of these instruments has demonstrated the ability of the majority of children, beginning in the latency period of development, to rate their own HRQoL. Increasingly, there has been a shift to relying on children as the *primary* respondents of their own HRQoL with recognition that there are situations in which self-report is not possible (age, illness, and impairment) [6, 8, 18, 22].

Despite the differences between parent and child ratings of the child's HRQoL, parent proxy reports do play a role in specific settings. A recent article by Upton et al. [19] presents two very different uses of parental ratings, one to enrich our understanding of the child's HRQoL by learning another perspective, and two, by using the parent report as a substitute for child report in instances where the child is unable or unavailable to provide self-report. In the first use, the concordance between raters is less important, whereas in the second, the degree of agreement between the raters would be of clinical importance.

In this study, we explore the relationship between child self-report and parent proxy report of HRQoL by domain, using the Pediatric Quality of Life Inventory™ (PedsQL™) 4.0 Generic Core Scales in a cohort of children newly diagnosed with cancer. Capitalizing on the rich database available through the previously conducted supportive care trial, involving 222 parent–child pairs, we describe the longitudinal nature of parent and child ratings of the child's HRQoL over the first 16 weeks following diagnosis. Specifically, we examine how age, intensity of treatment, and time on study affect the relationship of parent and child report for each of the domains within the measure. In addition to the expected variation in ratings by child age and domain, as previously reported, we hypothesized that concordance between parent and child would vary by treatment intensity and by time. Specifically, we anticipated that there would be better agreement among parent–child dyads receiving more intense therapy and that agreement would increase over time as parent and child spent more time together in a “shared reality.”

Patients and methods

Summary of original efficacy trial

A double-blind, placebo-controlled study was conducted at 27 sites in the United States to test the tolerability and effects of once-weekly intravenous epoetin alfa (EPO) on patient-reported HRQoL and clinical outcomes in anemic

pediatric patients receiving myelosuppressive treatment for cancer [17]. The study sample was stratified by tumor type (solid tumor/Hodgkin's disease or ALL/NHL). Patients were randomly assigned to treatment arms and were followed every 3–4 weeks (depending on their chemotherapy schedule) with a final visit at Week 16. Patients and their parents reported on the patients' HRQoL at baseline, Week 4 or 5, 7, 9 or 10, 13 and Week 16, using the age-appropriate forms of the PedsQL™ 4.0 Generic Core Scales and PedsQL™ 3.0 Cancer Module (patient-report only). The primary study endpoint was mean change from baseline to Week 16 in patient-reported PedsQL™ 4.0 Generic Core Scales. Enrollment goals to achieve 80% power at $\alpha = .05$ were achieved with 113 patients randomly assigned to receive EPO and 111 assigned to receive placebo. Patients did not differ at baseline demographically or in terms of their HRQoL scores; additionally, no significant differences in HRQoL were observed between the two groups at Week 16. Because HRQoL scores were similar between the two treatment groups, data from the two treatment arms were pooled for the analysis of observer agreement in the current study.

Study measure

The primary outcome measure is the age- and rater-specific version of the PedsQL™ 4.0, a multidimensional health profile measure of HRQoL. The PedsQL™ 4.0 contains 23 items forming four principal domains including physical functioning (8 items), emotional functioning (5 items), school functioning (5 items), and social functioning (5 items). The Acute Version of the PedsQL™ 4.0 has a reference period of 1 week and uses a Likert-type response scale for each item. Higher scores connote better functioning. Although summary subscales (physical and psychosocial functioning) and total scores can be generated, we elected to rely solely on the domain scores, based on the study's goals. This instrument has been shown to have acceptable internal consistency, known groups, and construct validity estimates, principally derived from cross-sectional samples [22–25]. There has been limited use of this instrument to assess child and parent agreement over time [7, 9].

Intensity of treatment

Each diagnosis and related treatment was reviewed by a pediatric oncologist and pediatric oncology nurse to establish criteria for designating levels of intensity of anti-cancer treatment. Initially, three levels of intensity (low, moderate, and high) were created based on patient diagnosis and stage, and type and duration of treatment. *Low intensity* included a short course of chemotherapy. *Mid-range intensity* included

treatment up to 1 year that was considered to be of moderate impact and toxicity. *High intensity* included multimodal therapy of greater than 6 months in duration, which was considered to be of greater intensity than the other two intensity levels. The two oncology specialists independently reviewed each diagnosis in addition to consulting with a pediatric oncologist at each of their respective settings. Consensus review followed, resulting in complete agreement between the two original pediatric oncology specialists.

Using these criteria, 12 patients were categorized as low intensity, 59 as mid-range intensity, and 148 as high intensity. Subsequently, for the purposes of analysis, the low and mid-range intensity categories were combined as non-high. In total, 66.7% of patients were categorized as high intensity and 33.3% of patients as non-high intensity.

Data completeness

The data set was examined for completeness. For missing items within each of the four domains of the PedsQL™ 4.0, guidelines established by the instrument's author [20], based on the “50% rule”, were used for imputation of the domain scores. While domain scores were missing for <1% of respondents (both raters) for the domains of physical, emotional, and social functioning (data not shown), overall missingness for school functioning exceeded 20% (both raters) at baseline. Given the high level of missingness within this domain, patterns of missingness within the school functioning domain were also examined by age group, treatment intensity, and visit (time interval).

Within-respondent measures

Internal consistency reliability was measured using the Cronbach's alpha [2] and examined for each domain of HRQoL by observer, age group, treatment intensity, and time interval. The criterion level of ≥ 0.70 was used to evaluate the adequacy of the internal consistency estimates [15].

Agreement between parent and child

To assess the extent of agreement between the child and parent assessments, we examined three components of the variation of each HRQoL domain using the model: $Y_{oit} = \mu_{oi} + \delta_{oi} + \varepsilon_{oit}$, where i indicates the i th child, o indicates whether the observer is the child (c) or parent (p), and t indicates the t th assessment. The first component assesses observer bias and is measured as difference between the child and parent group means averaged over the assessments ($\sum_{t=1}^T (\hat{\mu}_{ct} - \hat{\mu}_{pt})/T$). The second component

assesses the agreement between the child and parent assessment of the individual child status compared to the group average or the *between-subject variation*, δ_{oi} . This second component is measured by the correlation of the child and parent deviations, $\rho(\delta_{ci}, \delta_{pi})$. Strong correlations of the between-subject variation would indicate that both observers consistently identified the child as having higher or lower scores than the “average” child. The third component assesses the agreement with respect to *within-subject changes* between assessments, ε_{oit} , and is measured by correlation of the child and parent deviations, $\rho(\varepsilon_{cit}, \varepsilon_{pit})$. The *intraclass correlation* (ICC), a measure of the correlation of scores within observer within time, was also calculated. The ICC is the ratio of the between-subject variation to the total variance. Mean differences and correlations were estimated using multivariate mixed-effect models. Standard errors for correlations were estimated using 500 bootstrap samples. Although there are no established criteria for the interpretation of such correlations, correlations of <0.29 are generally considered low, 0.3–0.6 is considered moderate, and >0.6 is considered strong [12]. All analyses were conducted using SAS version 9.1. Additional details are presented in the “Appendix”.

Results

Study sample

All patients included in the efficacy analysis of the original trial ($n = 222$) were included in the analysis for the aims of this study. The majority of patients were between the ages of 13 and 18 years and Caucasian, and slightly more than half were boys (Table 1). There was a variety of diagnoses with acute lymphoblastic leukemia (ALL) being the most frequent. All subjects were on a regimen with a risk of chemotherapy-induced anemia with roughly two-thirds classified as having received a high-intensity regimen.

Missing assessments

For the domains of physical, emotional, and social functioning, low levels of missingness ($\leq 1\%$) were observed at the item level, indicating that domain scores did not require imputation for either rater group or across age groups (data not shown). In contrast, high rates of missingness were noted in the domain of school functioning (Table 2). Younger children had higher rates of missingness than older children. Across all age groups and both raters, the highest rates of missingness were detected at Week 7.

Table 1 Patient and disease characteristics ($n = 222$)

Variable	<i>N</i> (%)
Mean age at study entry	11.6 years (s.d. 3.9)
Age group	
5–7 years	47 (21.2)
8–12 years	67 (30.2)
13–18 years	108 (48.6)
Gender	
Male	121 (54.5)
Female	101 (45.5)
Race/ethnicity	
Caucasian	152 (68.5)
Black	27 (12.2)
Asian	3 (1.4)
Hispanic	27 (12.2)
Other	13 (5.1)
Diagnosis	
ALL	75 (33.8)
Lymphoma	
Hodgkin's	27 (12.2)
Non-Hodgkin's	22 (9.9)
Bone tumors	
Ewing's sarcoma	34 (15.3)
Osteosarcoma	22 (9.9)
Other peripheral solid tumors	
Rhabdomyosarcoma	15 (6.8)
Wilms' tumor	8 (3.6)
Neuroblastoma	2 (0.9)
Other	17 (7.7)
Treatment intensity	
High	148 (66.7)
Non-high	71 (32.0)
Missing	3 (1.3)
Mean baseline hemoglobin	9.65 g/dL (s.d. 1.16)

Internal consistency estimates

Internal consistency estimates varied by HRQoL domain, observer, and child age (Table 3). Overall, parental and child reliability was highest for physical functioning and lowest for social functioning. Among the child ratings of social functioning, internal consistency estimates were slightly below the criterion level of >0.70 across all age groups, treatment intensity, and time intervals. Similarly, the reliability estimates for emotional functioning and school functioning were approximately 0.6 for children in the 5- to 7-year age group. Otherwise, the domains for both child and parent observers meet the criterion for acceptable internal consistency reliability.

Between- and within-rater agreement

Children consistently reported higher functioning than their parents (Table 4), with the largest overall differences occurring in emotional functioning (9.1 points) and the smallest differences in physical functioning (2.8 points). Difference in social and school function were 5.0 and 5.4 points, respectively. Differences varied by age, with the biggest differences occurring in the oldest children (13–17 years) for emotional, social and school functioning and in the youngest children (5–7 years) for physical functioning. No differences were associated with the intensity of treatment. Differences between average child and parent reports of emotional and social functioning decreased as the study progressed (0–7 weeks vs. 9/10–16 weeks).

Agreement between the individual parent and child responses is summarized in Table 5. The correlation of the child and parent report of the between-subject variation for the entire group ranged from 0.61 (social functioning) to 0.86 (physical functioning). Within the subgroups defined

Table 2 Percentage missing within school functioning domain by rater, age group, and treatment intensity

Observer	Group	Week on trial					
		0	4/5	7	9/10	13	16
Parent	Overall	21.6	34.2	68.9	33.8	41.9	38.7
	Treatment intensity						
	Non-high intensity	23.0	33.8	60.8	40.5	44.6	40.5
	High intensity	20.9	34.5	73.0	30.4	40.5	37.8
Child	Overall	22.5	28.8	66.7	33.8	38.7	38.3
	Age group						
	5–7 years	34.0	38.3	70.7	36.2	48.9	40.4
	8–12 years	20.9	28.4	64.2	29.9	26.9	26.9
	13–18 years	18.5	25.0	66.7	35.2	41.7	44.4
	Treatment intensity						
Non-high intensity	24.3	31.1	58.1	41.9	47.3	44.6	
High intensity	21.6	27.7	70.9	29.7	34.5	35.1	

Table 3 Internal consistency (Cronbach's alpha)

Time	Group	Physical (8 items)	Emotional (5 items)	Social (5 items)	School (5 items)
Parent	Overall	0.91	0.84	0.71	0.82
Child	Overall	0.90	0.76	0.68	0.76
Age group	Age 5–7	0.82	0.59	0.61	0.60
	Age 8–12	0.92	0.82	0.69	0.74
	Age 13–17	0.90	0.80	0.70	0.81
Treatment intensity	Non-high intensity	0.87	0.79	0.66	0.74
	High intensity	0.91	0.75	0.68	0.76
Time interval	Week 0–7	0.89	0.76	0.67	0.78
	Week 9/10–16	0.91	0.76	0.68	0.73

Estimates in bold meet criterion standard of ≥ 0.70 . [Nunnally, 1994]

by age, treatment intensity, and time on study, the correlations were similar with two exceptions. The agreement of parent and child was significantly lower for the youngest children for the measures of physical and school function.

The correlation between child and parent in the within-subject variation in the entire group ranged from 0.27 (social function) to 0.37 (physical function). Again, the correlations are stable across age, treatment, and time, with the exception of lower scores for adolescents in measures of social and emotional functioning.

Finally, the correlation of measures within-respondent across time (ICC) was slightly higher for child reports (0.53–0.67) than parent reports (0.48–0.57). The ICCs were relatively constant across age, treatment, and time, with the exception of lower scores for the youngest children in measures of physical, social, and school functioning.

Discussion

Notable strengths of this work include the large number of child/parent dyads reporting serially about the child's HRQoL during the first 16 weeks of treatment, use of age-specific matched child/parent HRQoL instruments, and the multiple statistical strategies used to assess child/parent agreement within HRQoL domains. To our knowledge, this is among the few studies that have addressed child and parent agreement using the PedsQL™ 4.0 longitudinally in a large sample of children with newly diagnosed cancer.

The low rate of missingness of data and the high rate of completeness of child and parent reports increase the trustworthiness of findings with the exception of school functioning. Given higher rates of missingness in the school functioning domain for both child and parent, across child age groups and data assessment points, findings from this domain are interpreted with caution. The extent of missingness in this domain and the greater degree of

missingness within certain age groups and diagnoses may imply that this domain is especially sensitive to treatment intensity and timing within the treatment course. Alternative explanations, such as school vacations or delayed school enrollment, are also plausible reasons for missing data in this domain, as previously reported [21]. Thus, it is likely that this degree of missingness in the school functioning domain will occur in other studies involving children in active treatment for cancer who are of school age, particularly when the acute version of the instrument is used. This argues for the collection of information about reasons for missingness.

The multiple ways in which we examined child/parent HRQoL agreement converge in their results to indicate that parent proxy reports of their children in treatment for cancer are reasonable estimates of the child's reports.

The internal consistency estimates for the social functioning domain as reported by children across all age, treatment, and time subgroups are lower than needed to conclude that the reports are reliable. Similarly, we found suboptimum levels of reliability among young raters (5–7 years) in the domains of emotional and school functioning. This is an important issue when examining agreement between raters because agreement at the individual level is dependent upon the reliability of the instrument. The low internal consistency estimates in these domains likely contribute to the observed lower child/parent agreement, as compared to other domains. Of note, for the more concrete and observable domain of physical functioning, high reliability was noted across all subgroups, as expected, with high correlation between observers and small mean differences.

The relationship between ICC and low reliability was formally addressed by Verrips et al. [26] who demonstrated methods for correcting ICC by dividing the observed ICC by the square root of the product of the raters' reliability. The authors suggest that correction for attenuation can

Table 4 Mean difference between parent and child

Measure	Group	Child's score		Parent's score		Difference		<i>t</i>	<i>P</i> value
		Mean	(SE)	Mean	(SE)	Mean	(SE)		
Physical	Overall	78.50	1.07	75.65	1.01	2.85	0.68	4.17	<0.0001
	Age 5–7	87.54	1.40	79.75	1.50	7.79	1.54	5.06	<0.0001
	Age 8–12	76.36	2.17	75.15	2.07	1.21	1.22	0.99	0.3214
	Age 13–17	75.91	1.49	74.19	1.46	1.72	0.89	1.92	0.0546
	Non-high intensity	82.66	1.50	80.17	1.43	2.49	1.23	2.02	0.0439
	High intensity	76.44	1.38	73.33	1.29	3.12	0.82	3.78	0.0002
	Weeks 0–7	77.73	1.12	74.41	1.12	3.33	0.91	3.67	0.0003
	Weeks 9/10–16	79.67	1.21	77.18	1.14	2.50	0.78	3.18	0.0015
Emotional	Overall	78.20	1.03	69.11	1.08	9.08	0.85	10.68	<0.0001
	Age 5–7	76.61	2.08	69.07	2.28	7.54	1.85	4.06	<0.0001
	Age 8–12	78.65	2.00	70.04	1.98	8.61	1.45	5.93	<0.0001
	Age 13–17	78.51	1.45	68.44	1.55	10.07	1.26	8.02	<0.0001
	Non-high intensity	78.70	1.85	70.36	1.87	8.34	1.47	5.66	<0.0001
	High intensity	78.22	1.23	68.56	1.32	9.66	1.05	9.22	<0.0001
	Weeks 0–7	76.60	1.15	66.22	1.22	10.37	1.08	9.64	<0.0001
	Weeks 9/10–16	80.60	1.11	72.74	1.17	7.89	0.97	8.07	<0.0001
Social function	Overall	83.51	0.85	78.52	0.93	4.99	0.85	5.88	<0.0001
	Age 5–7	80.19	1.81	79.13	1.71	1.05	1.68	0.63	0.5302
	Age 8–12	83.16	1.67	79.89	1.58	3.27	1.38	2.37	0.0180
	Age 13–17	85.25	1.09	77.43	1.46	7.81	1.26	6.18	<.0001
	Non-high intensity	84.98	1.35	79.66	1.81	5.32	1.59	3.34	0.0009
	High intensity	82.92	1.07	77.94	1.05	4.98	1.00	4.99	<0.0001
	Weeks 0–7	83.12	0.93	76.96	1.05	6.17	1.03	6.01	<0.0001
	Weeks 9/10–16	84.25	0.94	80.55	1.05	3.70	0.91	4.05	<0.0001
School function	Overall	71.43	1.20	66.98	1.34	4.44	1.05	4.22	<0.0001
	Age 5–7	71.59	2.20	67.26	2.55	4.33	2.75	1.57	0.1167
	Age 8–12	73.02	2.02	70.72	2.22	2.30	1.73	1.33	0.1844
	Age 13–17	70.49	1.90	64.54	2.09	5.96	1.49	3.99	<0.0001
	Non-high intensity	70.23	2.14	66.16	2.18	4.08	1.60	2.55	0.0111
	High intensity	72.21	1.45	67.61	1.69	4.61	1.34	3.43	0.0006
	Weeks 0–7	69.76	1.37	65.69	1.58	4.07	1.39	2.93	0.0036
	Weeks 9/10–16	73.41	1.42	68.73	1.56	4.68	1.27	3.69	0.0002

Note that statistics are tests for within-group differences, not that the difference varies across the groups

raise ICC by 10–15%. While we do not present the results of corrected ICC for our study population, we noted similar magnitude of the corrections.

As reported in previous studies involving children with different illnesses and their parents reporting on the child's HRQoL [6, 27, 28], children in our study consistently reported higher HRQoL functioning than did their parents, especially in the domain of emotional functioning. We attribute this consistent difference to a rating bias. Parent ratings of HRQoL reported in other studies indicate that parents over or underestimate their child's HRQoL compared to the child's ratings, with particular discrepancies

reported in emotion-focused items [5, 6, 10, 11, 14, 28]. The rating differences in this domain appear to become more discordant with adolescents as compared to younger children [1]. Our findings also include discrepancies in social and school functioning ratings between young children (ages 5–7 years) and parents. In addition to the potential contribution of lower reliability among these young raters, a recently proposed explanation of differences between child and parent HRQoL ratings is that children tend to base their HRQoL reports on a recent event, whereas parents tend to base their reports on multiple events [4]. It is likely that information about the

Table 5 Between and within parent and child correlations

Measure	Group	Agreement between parent and child				Correlation of responses within respondent			
		Between-subject variation		Within-subject variation		Within child		Within parent	
		Corr	(SE)	Corr	(SE)	ICC	(SE)	ICC	(SE)
Physical	Overall	0.86	0.03	0.37	0.06	0.67	0.03	0.52	0.04
	Age 5–7	0.49	0.20	0.37	0.11	0.38	0.10	0.22	0.08
	Age 8–12	0.90	0.05	0.32	0.11	0.72	0.05	0.59	0.06
	Age 13–17	0.89	0.04	0.39	0.07	0.66	0.05	0.57	0.05
	Non-high intensity	0.72	0.08	0.30	0.14	0.59	0.06	0.48	0.07
	High intensity	0.89	0.03	0.40	0.05	0.68	0.04	0.51	0.04
	Weeks 0–7	0.84	0.06	0.32	0.09	0.60	0.05	0.47	0.06
Emotional	Weeks 9/10–16	0.84	0.04	0.39	0.07	0.81	0.03	0.66	0.05
	Overall	0.74	0.04	0.30	0.04	0.58	0.04	0.57	0.03
	Age 5–7	0.71	0.13	0.24	0.09	0.52	0.09	0.61	0.08
	Age 8–12	0.79	0.08	0.42	0.06	0.56	0.08	0.56	0.05
	Age 13–17	0.73	0.06	0.22	0.05	0.64	0.05	0.57	0.05
	Non-high intensity	0.74	0.08	0.25	0.06	0.66	0.06	0.60	0.05
	High intensity	0.74	0.06	0.33	0.04	0.54	0.05	0.55	0.05
Social function	Weeks 0–7	0.69	0.06	0.34	0.06	0.59	0.05	0.57	0.05
	Weeks 9/10–16	0.72	0.05	0.29	0.05	0.61	0.05	0.66	0.04
	Overall	0.61	0.07	0.27	0.05	0.57	0.03	0.52	0.04
	Age 5–7	0.61	0.13	0.37	0.10	0.45	0.08	0.39	0.07
	Age 8–12	0.72	0.12	0.31	0.06	0.63	0.05	0.49	0.06
	Age 13–17	0.61	0.09	0.18	0.08	0.57	0.05	0.58	0.06
	Non-high intensity	0.59	0.10	0.20	0.08	0.57	0.06	0.60	0.06
School function	High intensity	0.63	0.09	0.31	0.05	0.56	0.04	0.46	0.05
	Weeks 0–7	0.62	0.09	0.21	0.08	0.52	0.05	0.48	0.06
	Weeks 9/10–16	0.67	0.07	0.20	0.07	0.67	0.04	0.65	0.04
	Overall	0.77	0.05	0.36	0.05	0.53	0.04	0.49	0.04
	Age 5–7	0.33	0.29	0.41	0.11	0.34	0.11	0.41	0.08
	Age 8–12	0.75	0.10	0.33	0.11	0.54	0.09	0.53	0.07
	Age 13–17	0.85	0.06	0.37	0.06	0.58	0.05	0.48	0.06
School function	Non-high intensity	0.85	0.07	0.43	0.08	0.61	0.06	0.38	0.08
	High intensity	0.74	0.07	0.34	0.06	0.49	0.05	0.54	0.05
	Weeks 0–7	0.71	0.08	0.35	0.09	0.57	0.06	0.48	0.06
	Weeks 9/10–16	0.75	0.06	0.32	0.09	0.63	0.06	0.60	0.06
	Overall	0.77	0.05	0.36	0.05	0.53	0.04	0.49	0.04

Note that statistics are tests for within-group differences, not that the difference varies across the groups

child's clinical status during treatment, often shared preferentially with the parent and not directly with the child, influences parent HRQoL ratings. In a previous study, the parent HRQoL ratings tended to be more highly associated with clinical indicators of the child's condition [3, 13, 16], a finding that has led to the suggestion that parent reports may be more accurate (in terms of disease status) than the child's self-reported HRQoL ratings. However, while HRQoL measures should be correlated with disease status, they are not a measure of disease status, and these data do not confirm that child's self-report is a less accurate measure of the impact of the disease and its treatment.

Importantly, our findings indicate that the size of the difference between child/parent ratings decreases over time and assessment points in three of the domains. This may indicate that a training effect is reflected in the serial completion of HRQoL assessments. Alternatively, the increasing alignment of the child and parent ratings suggests that the experience of the cancer treatment becomes a more shared experience between child and parent over time. From a measurement perspective, the latter interpretation supports the role of a parent proxy in reporting on their child's HRQoL during treatment for cancer when the child is unable to report. There are important developmental

implications to these findings, particularly in adolescence, when creation of separate identity and perception is the norm. The illness experience may thrust the dyad into increased closeness (Eiser, personal communication, 1998).

Previous research studies emphasizing the differences between child and parent reports were cross-sectional or measured HRQoL ratings at a single time point. Our findings of decreasing differences in agreement indicate that conclusions about child/parent HRQoL agreement should not be based on a single measurement point or even on the first of a series of measurement points. Moreover, our findings point to the need for further research about parent–child agreement over time, extending beyond the acute treatment phase.

Conclusion

The moderate-to-strong agreement between child and parent ratings, particularly in the domains of physical and emotional functioning, supports the usefulness of parent proxy reports during a child's initial treatment for cancer. Interestingly, the agreement improved in these two domains quickly (by Week 9/10 of therapy). Although agreement levels may not support replacing child reports with parent reports, they do indicate the value of the parent proxy reports as ancillary data that help to interpret the child's illness experience. We recommend collecting HRQoL data from both the child and the parent proxy across the course of treatment for a child with a life-threatening illness.

Acknowledgments The authors express sincere appreciation to the patients, their parents, and the investigators of the parent trial; Drs. Bassem I. Razzouk, Jeffrey D. Hord, Marilyn Hockenberry, James Feusner, Denise Williams, and Wayne R. Rackoff; as well to Helen Varsos, MS, RPh, Ortho Biotech Clinical Affairs LLC. We also appreciate the permission from Ortho Biotech Clinical Affairs LLC for granting us access to the data set used for this study. This work was funded in part by Ortho Biotech.

Appendix

The longitudinal data for each domain were modeled as follows: $Y_{oit} = \mu_{ot} + \delta_{oi} + \varepsilon_{oit}$ where i indicates the i th child, o indicates whether the observer is the child (c) or parent (p), and t indicates the t th assessment. The between-subject random effects and within-subject residual errors were allowed to be correlated between the child and parent, $\text{Cov}(\delta_{ci}, \delta_{pi}) = \zeta_{cp}$ and $\text{Cov}(\varepsilon_{cit}, \varepsilon_{pit}) = \sigma_{cp}$ but otherwise the errors were independent, $\text{Cov}(\delta_{oi}, \varepsilon_{oit}) = 0$.

The SAS code to generate the means and variance parameters is as follows:

```
proc mixed data=work.mixed method=ml covtest;
  class pid observer visit;
  model (outcome_specified_here)=observer*visit/noint
solution;
  random observer/subject=pid type=un;
  repeated observer/subject=pid (visit) type=un;
  where (specific_for_subgroups);
run;
```

This will generate 6 variance parameters that correspond to the output as:

ζ_c^2	UN (1,1)	PID
ζ_{cp}	UN (2,1)	PID
ζ_p^2	UN (2,2)	PID
σ_c^2	UN (1,1)	PID (visit)
σ_{cp}	UN (2,1)	PID (visit)
σ_p^2	UN (2,2)	PID (visit)

The correlation of the two between-observer components are calculated as $\zeta_{cp}/\sqrt{\zeta_c^2\zeta_p^2}$ for between-subject variation and $\sigma_{cp}/\sqrt{\sigma_c^2\sigma_p^2}$ for within-subject variation.

The interclass correlations are calculated as $\zeta_o^2/(\zeta_o^2 + \sigma_o^2)$.

References

1. Chang, P. C., & Yeh, C. H. (2004). Agreement between child self-report and parent proxy-report to evaluate quality of life in children with cancer. *Psychooncology*, *14*, 125–134.
2. Cronbach, L. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, *16*, 297–334. doi:10.1007/BF02310555.
3. Czystewski, D. I., Mariotto, M. J., Bartholomew, L. K., LeCompte, S. H., & Sockrider, M. M. (1994). Measurement of quality of well being in a child and adolescent cystic fibrosis population. *Medical Care*, *32*, 965–972.
4. Davis, E., Nicolas, C., Waters, E., Cook, K., Gibbs, L., Gosch, A., et al. (2007). Parent-proxy and child self-reported health-related quality of life: Using qualitative methods to explain the discordance. *Quality of Life Research*, *16*, 863–871.
5. Eiser, C., & Eiser, J. R. (2000). Social comparisons and quality of life among survivors of childhood cancer and their mothers. *Psychology & Health*, *15*, 435–450.
6. Gong, G. W. K., Young, N. L., Dempster, H., Porepa, M., & Feldman, B. M. (2007). The quality of my life questionnaire: The minimal clinically important difference for pediatric rheumatology patients. *Journal of Rheumatology*, *34*, 581.
7. Hinds, P. S., Billups, C. A., Cao, X., Gattuso, J. S., Burghen, E., West, N., et al. (2009). Health-related quality of life in adolescents at the time of diagnosis with osteosarcoma or acute myeloid leukemia. *European Journal of Oncology Nursing*, *13*, 156–163.

8. Hinds, P. S., Brandon, J., Allen, C., Hijjiya, N., Newsome, R., & Kane, J. R. (2007). Patient-reported outcomes in end-of-life research in pediatric oncology. *Journal of Pediatric Psychology*, *32*, 1079–1088.
9. Hinds, P. S., Gattuso, J. S., Billups, C. A., West, N. K., Wu, J., Rivera, C., et al. (2009). Aggressive treatment of non-metastatic osteosarcoma improves health-related quality of life in children and adolescents. *European Journal of Cancer*, *45*, 2007–2014.
10. Klassen, A. F., Miller, A., & Fine, S. (2006). Agreement between parent and child report of quality of life in children with attention-deficit/hyperactivity disorder. *Child: Care, Health and Development*, *32*, 397–406.
11. Koopman, H. M., Koetsier, J. A., Taminiu, A. H. M., Hijnen, K. E., Bresters, D., & Egeler, R. M. (2005). Health-related quality of life and coping strategies of children after treatment of a malignant bone tumor: A 5-year follow-up study. *Pediatric Blood & Cancer*, *45*, 694–699.
12. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310.
13. Lustig, R. H., Hinds, P. S., Ringwald-Smith, K., Christensen, R. K., Kaste, S. C., Schreiber, R. E., et al. (2003). Octreotide therapy of pediatric hypothalamic obesity: A double-blind, placebo-controlled trial. *Journal of Clinical Endocrinology and Metabolism*, *88*, 2586–2592.
14. Noyes, J. (2007). Comparison of ventilator-dependent child reports of health-related quality of life with parent reports and normative populations. *Journal of Advanced Nursing*, *58*, 1–10.
15. Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill, Inc.
16. Parsons, S., Terrin, N., Ratichek, S., Tighiouart, H., & Chang G., For the Journeys to Recovery Study. (2007). The influence of clinical events after HSCT on parent and child assessments of HRQL. International society for quality of life research meeting abstracts. *The QLR Journal A*, *38*, 1371.
17. Razzouk, B. I., Hord, J. D., Hockenberry, M., Hinds, P. S., Feusner, J., Williams, D., et al. (2006). Double-blind, placebo-controlled study of quality of life, hematologic end points, and safety of weekly epoetin alfa in children with cancer receiving myelosuppressive chemotherapy. *Journal of Clinical Oncology*, *24*, 3583–3589.
18. Sherifali, D., & Pinelli, J. (2007). Parent as proxy reporting: Implications and recommendations for quality of life research. *Journal of Family Nursing*, *13*, 83–98.
19. Upton, P., Lawford, J., & Eiser, C. (2008). Parent-child agreement across child health-related quality of life instruments: A review of the literature. *Quality of Life Research*, *17*, 895–913.
20. Varni, J. W. (2010). *The PedsQL scoring algorithm: Scoring the pediatric quality of life inventory*. Retrieved April 16, 2007 from <http://www.pedsq.org/score.html>.
21. Varni, J. W., Burwinkle, T. M., Katz, E. R., Meeske, K., & Dickinson, P. (2002). 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*, *94*, 2090–2106.
22. Varni, J. W., Lane, M. M., Burwinkle, T. M., Fontaine, E. N., Youssef, N. N., Schwimmer, J. B., et al. (2006). Health-related quality of life in pediatric patients with irritable bowel syndrome: A comparative analysis. *Journal of Developmental and Behavioral Pediatrics*, *27*, 451–458.
23. Varni, J. W., Limbers, C., & Burwinkle, T. (2007). How young can children reliably and validly self-report their health-related quality of life?: An analysis of 8,591 children across age subgroups with the PedsQL™ 4.0 Generic Core Scales. *Health and Quality of Life Outcomes*, *5*, 1.
24. Varni, J. W., Limbers, C. A., & Burwinkle, T. M. (2007). Parent proxy-report of their children's health-related quality of life: An analysis of 13, 878 parents' reliability and validity across age subgroups using the PedsQL™ 4.0 Generic Core Scales. *Health and Quality of Life Outcomes*, *5*, 1–10.
25. Varni, J. W., Seid, M., Knight, T. S., Uzark, K., & Szer, I. S. (2002). The PedsQL™ 4.0 Generic Core Scales: Sensitivity, responsiveness, and impact on clinical decision-making. *Journal of Behavioral Medicine*, *25*, 175–193.
26. Verrrips, G. H., Vogels, A. G., den Ouden, A. L., Paneth, N., & Verloove-Vanhorick, S. P. (2000). Measuring health-related quality of life in adolescents: Agreement between raters and between methods of administration. *Child: Care, Health and Development*, *26*, 457–469. doi:10.1046/j.1365-2214.2000.00181.x.
27. Verrrips, G. H., Vogels, T. G. C., Koopman, H. M., Theunissen, N. C. M., Kamphuis, R. P., Fekkes, M., et al. (1999). Measuring health-related quality of life in a child population. *The European Journal of Public Health*, *9*, 188–193.
28. Youssef, N. N., Murphy, T. G., Langseder, A. L., & Rosh, J. R. (2006). Quality of life for children with functional abdominal pain: A comparison study of patients' and parents' perceptions. *Pediatrics*, *117*, 54–59.