

---

ORIGINAL ARTICLES

# Psychometric properties of the EORTC Quality of Life Questionnaire in inpatient cancer rehabilitation in Germany

---

JÖRG DIRMAIER, M.D.,<sup>1</sup> SILKE ZAUN, PH.D., M.D.,<sup>2</sup> UWE KOCH,<sup>1</sup> TIMO HARFST,<sup>1</sup>  
AND HOLGER SCHULZ, PH.D.<sup>1</sup>

<sup>1</sup>Institute and Polyclinic for Medical Psychology, Centre for Psychosocial Medicine, University Clinic  
Hamburg-Eppendorf, Hamburg, Germany

<sup>2</sup>Nordfriesland-Clinic, St. Peter-Ording, Germany

(RECEIVED February 13, 2004; ACCEPTED May 2, 2004)

## ABSTRACT

*Objective:* Recent years have shown an increase in the use of questionnaires measuring health-related quality of life to verify the quality of treatment in the field of oncology. An often used cancer-specific questionnaire is the “Quality of Life Core Questionnaire of the European Organization for the Research and Treatment of Cancer” (EORTC QLQ-C30). The purpose of this study is to analyze the psychometric properties of the EORTC QLQ-C30 (version 1) in order to determine the feasibility and appropriateness for its use in inpatient cancer rehabilitation in Germany with heterogeneous diagnoses.

*Methods:* The questionnaire was administered to a sample of 972 cancer patients at the beginning of treatment and to 892 patients after treatment. Besides descriptive analysis, the statistical analyses include confirmatory analysis and the multitrait/multimethod approach to test the questionnaire’s postulated scale structure (factorial validity) and its reliability (internal consistencies). The analysis also includes a comparison of responsiveness indices (effect size, reliable change index) to test the sensitivity of the instrument.

*Results:* The EORTC QLQ-C30 showed satisfactory levels of reliability and sensitivity, but the postulated scale structure could not be confirmed. The results illustrate that the varimax-rotated solution of a principal component analysis does not confirm the scale structure postulated by the authors. Correspondingly, the selected fit indices within the scope of the confirmatory factor analysis do not show satisfactory results either.

*Significance of results:* We therefore consider version 1 of the EORTC QLQ-C30 to be only limitedly useful for the routine assessment of changes in the quality of life of cancer patients in inpatient rehabilitation in Germany, especially because of the instrument’s length and possible redundancies. For this reason, a scoring procedure limited to a subset of items is suggested, revealing satisfactory to good psychometric indices. However, further psychometric tests are necessary, especially with regard to validity and sensitivity.

**KEYWORDS:** Cancer, Quality of life, Rehabilitation, Sensitivity, Validity

## INTRODUCTION

As a result of rising costs within the public health care system, increasing efforts have been made in the past to distribute the available funds according

to principles of quality assurance. Consequently, in the fields of oncological rehabilitation and palliative care, rising endeavors have been made to verify and optimize the quality of care (Krischke & Petermann, 1995; Stump et al., 1998; Küchler et al., 1999). Nevertheless, due to only limited practical use of common outcome criteria of hospitalized cancer treatment, such as survival time and tumor response, within the scope of rehabilitative and palliative treatment of oncological diseases, stan-

---

Corresponding author: Jörg Dirmaier, University Clinic Hamburg-Eppendorf, Centre for Psychosocial Medicine, Institute and Polyclinic for Medical Psychology, House S 35, 20246 Hamburg, Germany. E-mail: dirmaier@uke.uni-hamburg.de

standardized outcome criteria have only hesitantly been developed. However, more recently, the patient's subjective perception of the quality of survival time was given more consideration, particularly in the palliative setting. Quality of life assessment is particularly relevant to patients with progressive conditions, particularly in the later phases of the disease (Morgan, 2000; Kyriaki et al., 2001; Paci et al., 2001; Kaasa & Loge, 2003). It is assumed that, independent of prognosis and the course of the illness, cancer influences subjective parameters of quality of life, for example, the perceived state of health or everyday activities, which can be measured by directly posing questions to the patient (Aaronson et al., 1988). The concept of health-related quality of life is therefore increasingly used to evaluate the course and outcome of oncological treatment and surveys that not only measure the strain that accompanies symptoms but also the changes in quality of life during the course of treatment gained high relevance in the field of rehabilitative and palliative care of oncological diseases (Osoba et al., 1994, 1998; King, 1996; Ringdal & Ringdal, 2000; Weis et al., 2000).

Extensive scientific studies have been carried out repeatedly to evaluate the effectiveness of oncological treatment with respect to the patients' quality of life (Schulz et al., 2001), providing some evidence of improvements in the quality of life directly following the end of treatment. An often used instrument to measure changes in the patients' subjective quality of life is the tumor-specific "Quality of Life Core Questionnaire of the European Organization for Research and Treatment of Cancer" (EORTC QLQ-C30; Aaronson et al., 1993). This questionnaire consists of 30 items that are arranged within six scales of functioning oriented aspects (physical functioning [PF], role functioning [RF], cognitive functioning [CF], emotional functioning [EF], social functioning [SF], and global quality of life [GQL]). In addition, the questionnaire includes three symptom scales (fatigue [F], nausea/vomiting [N/V], and pain [P]) as well as six symptom items (dyspnoea, loss of appetite, sleep disorder, constipation, diarrhea, and financial problems). These different dimensions are measured in a period of one week. In version 1 used in this study, the patients answered the questions either dichotomously with a "yes/no" response (items 1–7) or on a 4-point Likert scale (1 = "not at all," 2 = "somewhat," 3 = "moderate," 4 = "very much") and a 7-point Likert scale (1 = "very bad" to 7 = "excellent"). In the current version 3.0, items 1–7 from the scales "physical functioning" and "role functioning" are measured on a 4-point Likert scale as well. Additionally, items 6 and 7 have been marginally

changed linguistically to adapt them to the altered scaling. The scores of the six functioning scales are calculated by first adding up the raw scores of a scale, then dividing by the number of items, and finally mapping the values for each parameter onto a scale of 0 to 100, with 100 representing the highest level of functioning. The symptom scales and items are calculated in a similar manner, except for the difference that higher scores mean a higher symptom burden. On average, it takes 11 min to complete the questionnaire (Aaronson et al., 1993). The questionnaire was validated in an international study with patients suffering from lung cancer (Aaronson et al., 1993) indicating that five (PF, RF, CF, SF, N/V) of the nine scales had reliability coefficients below 0.7, two of them (RF, CF) below 0.6. In addition, three subscales (PF, RF, F) showed substantial interscale correlations. Currently, the EORTC Quality of Life Group is actively developing a shortened version of the questionnaire.

So far, there are only a few examinations of the psychometric properties of the EORTC QLQ-C30 in the field of inpatient oncological rehabilitation in Germany in which patients receive both active anti-tumor treatment, such as chemotherapy, as well as supportive care. Krischke and Petermann (1995) came to the conclusion that the instrument's psychometric properties are insufficient with relationship to the scales, which show only modest internal consistencies and high floor effects.

Because version 1 is still widely used in rehabilitative and palliative care of oncological diseases in Germany (Küchler et al., 1999; Weis et al., 2000), it seems appropriate to test the psychometric properties of the questionnaire once again. Such an analysis should be conducted with particular regard to the use of the questionnaire with a more extensive sample of heterogeneous cancer diagnoses and the need for reference data within these kind of studies. Due to the instrument's intended use in the measurement of changes in quality of life, not only factorial validity but also sensitivity to change is to be the focus of examination. When appropriate, because of the instrument's length and possible redundancies, an abridged scoring procedure is to be developed that possibly surveys the relevant fields of symptoms in oncological rehabilitation more economically and clearly.

## METHODS

### Sample

A consecutive sample of patients from the Nordfriesland Clinic in St. Peter-Ording, Germany, and

the clinical oncology section of the Habichtswald Clinic in Kassel-Wilhelmshöhe, Germany, was questioned. The clinics offer a rehabilitative treatment program, focusing on the reduction of symptoms and the improvement of the ability to cope with the resulting changes and limitations due to the disease. Possible indications include all solid as well as malignant hematological tumors. The basic medical treatment consists of all necessary internal oncological as well as chemotherapeutical measures. Additionally, overall therapeutic measures such as, for example, physical therapy, pain therapy, or training in coping skills are performed. To measure quality of life, the EORTC QLQ-C30 (version 1) was administered at the time of admission and at discharge. The sample contained 972 patients at the beginning and 892 at the end of treatment. The patients were, on average, 53 years old; 75% of the patients were women. The patients suffered from various tumors, with breast cancer (48%) being the most frequent. Eighty-four percent of the patients had cancer for the first time (see Table 1). The distribution of the socio-demographic and clinical data essentially correspond to that of other oncological rehabilitation clinics in Germany (Küchler et al., 1999; Weis et al., 2000).

**Table 1.** Clinical characteristics of the sample ( $N = 972$ )

	Patients	
	N <sup>a</sup>	%
Age (years, mean $\pm$ SD)	52 $\pm$ 11.4	
Gender		
Female	625	76%
Male	202	24%
Diagnoses		
Breast cancer	369	48%
Intestinal cancer	86	11%
Gynecological cancer	69	9%
Lymph cancer	55	7%
Prostate cancer	46	6%
Lung cancer	39	5%
Other diagnoses	79	10%
Type of disease		
Primary disease	563	84%
Relapse	52	8%
Multiple diseases	59	9%

<sup>a</sup>The numbers do not add up to the totals due to missing data.

## Statistical Analyses

### Testing the Psychometric Properties

The factorial structure of the questionnaire was tested by means of a principal components analysis, a confirmatory factor analysis, and the multi-trait/multimethod approach of Campbell and Fiske (1959). In addition, the psychometric testing involves descriptive item analysis and the calculation of internal consistencies.

The principal components analysis was performed by determining the factors in accordance with the guidelines of the authors (Aaronson et al., 1993). The results of this analysis were then compared with the varimax-rotated solution based on the current data set with regard to the assignment of the items to the respective scales.

For a confirmatory testing of the postulated scale structure and to examine whether the factor structure can be confirmed with this sample of patients of oncological rehabilitation with heterogeneous diagnoses, the program AMOS 4.0 (Arbuckle & Wothke, 1999) was used, extracting a factor model that is being tested regarding its suitability. The model's suitability was determined by use of the "Comparative Fit Index" (CFI) and the "Non-Normed-Fit Index" (NNFI), regarding values  $>0.90$  as acceptable (Hu & Bentler, 1999). The Chi<sup>2</sup>-Test (corrected for degrees of freedom) was not used because it is much more conservative within confirmatory analyses of large samples (Hu & Bentler, 1999). As a further measurement for assessing the sufficiency of fit, the "Root Mean Square Error of Approximation" (RMSEA) was also calculated. According to Browne and Cudeck (1993), in the interpretation of the RMSEA results, values  $<0.05$  are considered to indicate a good fit (model confirmed), values between 0.05 and 0.08 ( $0.05 \leq \text{RMSEA} < 0.08$ ) indicate a moderate fit, and values of  $\text{RMSEA} \geq 0.08$  indicate a poor fit.

A multitrait/multimethod approach was conducted by the MAP program (Hays, 1991) to ascertain item-scale correlations as well as convergent and discriminant validity. Item-convergent validity is assessed by determining whether each item in a scale is substantially related to the total score computed from other items in that scale. Item internal consistency is supported if an item correlates substantially ( $r > 0.40$ ) with the scale it is hypothesized to represent (Hays, 1991). To correct for overlap, the hypothesized item is deleted from the scale with which it is correlated. Determining that each item correlates most strongly with the scale to which it is hypothesized to belong assesses the item's discriminant validity criterion. Item-

discriminant validity is dependent on the magnitude of the correlation between an item and its scale relative to the correlation of that item to other scales. A scale fit is computed indicating the percentage of those items that correlate significantly higher to their own scale than to each of the other scales. This scale fit can take on values between 0% and 100% and indicates the factorial validity of the questionnaire on an item level. Values higher than 90% are considered to confirm the postulated scale (Hays, 1991). All 24 items used by Aaronson et al. (1993) for constructing the scales were included in the analysis.

### *Sensitivity*

Recently, research in quality of life has increasingly analyzed the ability of an instrument to measure changes over time (Wiebe et al., 1997; Pfenning et al., 1999; Liang, 2000). Because widely accepted criteria or procedures have not yet been developed (Liang, 1995, 2000), basic methods mentioned in the literature are compared.

First, the correlation coefficients of the scales between the two points of measurement are calculated. Furthermore, to be able to determine significant differences in means, *t* tests for paired samples at a significance level of 5% were used. Because tests of significance depend on the sample size, effect sizes (*d*) were calculated in accordance with Cohen (1988). The coefficients can be distinguished in categories of small ( $0.2 \leq d < 0.5$ ), moderate ( $0.5 \leq d < 0.8$ ), and large ( $d > 0.8$ ) effect sizes. However, effect size measurements have the restriction of being of only limited use for slowly progressing diseases because they use the mean change as the numerator (Pfenning et al., 1999). Another disadvantage is that measures of effect size assume that all patients change in the same direction (Liang, 2000). Therefore, the "Reliable Change Index" (RCI) was calculated, representing another measurement of sensitivity (Jacobson & Truax, 1991). The RCI is computed by dividing the difference between the pre- and posttreatment scores by the standard error of the difference between the two scores, which is composed of the standard deviation of the measure and the reliability of the instrument. If the product is larger than the *z*-score desired level of significance, in this case 1.96 ( $p \leq 0.05$ ), the change in pre-to-posttreatment scores is said to occur beyond that of chance variation.

Because up-to-date, norm values for patients with heterogeneous diagnoses in Germany do not exist, and considering the substantial size of the sample, effect sizes and the RCI were calculated by using the corresponding standard deviation and reliabil-

ity coefficients at the beginning of the treatment derived from the present sample. Because the dispersion of the sample at hand shows rather conservative estimates in comparison with values of an extensive international study (Ringdal & Ringdal, 1993), this procedure seemed to be appropriate.

## RESULTS

### Factorial Validity

A descriptive analysis of the missing values of all items showed a mean rate of 1.6%; the items varied between 0.2% and 2.5% missing values. For the majority of the analyses, missing values were replaced according to recommendations of Peng et al. (Priv Comm) by means of the expectation-maximization method. However, missing values were not replaced but omitted when calculating the confirmatory factor analysis in order not to affect the results.

The principal component analysis (with replacement of missing values;  $n = 972$ ) shows a distribution of eigenvalues as presented (8.2, 2.0, 1.5, 1.4, 1.1, 1.0, 0.9, 0.9, 0.8, 0.7; indicating only the first 10 values); the extracted nine factors explain 74% of variance. The varimax-rotated solution does not confirm the scale structure postulated by the authors. Correspondingly, the selected fit indices within the scope of the confirmatory factor analysis used (without replacement of missing values,  $n = 755$ ) also show unsatisfactory results (CFI = 0.85; NNFI = 0.82, RMSEA = 0.08). As can be seen in Table 2, the calculation of the correlations between the scales before the treatment shows moderate to high values ( $0.30 \leq r \leq 0.65$ ). Lower correlations with the other scales can be shown for the scale "nausea/vomiting" ( $0.17 \leq r \leq 0.31$ ). The calculation of the mean value of all correlation coefficients (after calculating the Fisher *Z*-transformed correlations, without the scale "global quality of life") results in a value of  $r_{\text{mean}} = 0.38$ .

Before treatment, the "scale fit" reaches 100% in six of the nine scales; the scales "physical functioning," "role functioning," and "emotional functioning" stay below 90% (see Table 3).

### Floor/Ceiling Effects

Because floor and ceiling effects (percentage of the lowest and highest possible scale value, respectively) depend on, among other things, the range of the scale, there is some indication for floor effects in the scales with dichotomous items. Prior to treatment, this concerns the scales "physical functioning" and "role functioning." Moderate floor effects



**Table 2.** Correlations between EORTC-QLQ-C30 scales

Scale (items)	Physical	Role	Cognitive	Emotional	Social	Fatigue	Na/Vo
<b>Functioning scales</b>							
Physical (1–5)							
Role (6, 7)	0.52						
Cognitive (21–24)	0.23	0.25					
Emotional (20, 25)	0.24	0.30	0.47				
Social (26, 27)	0.40	0.42	0.40	0.52			
<b>Symptom scales</b>							
Fatigue (10, 12, 18)	0.53	0.46	0.45	0.56	0.53		
Nausea/vomiting (14, 15)	0.20	0.17	0.19	0.25	0.24	0.31	
Pain (9, 19)	0.40	0.36	0.23	0.35	0.35	0.45	0.29

prior to treatment can be found despite a 4-point scale in the scales “cognitive functioning” and “pain”; high floor effects are shown for the scale “nausea/vomiting” (see Table 3).

### Reliability

At the beginning of the treatment, the calculated reliability coefficients (Cronbach’s  $\alpha$ ) meet the minimum requirements of  $\alpha \geq 0.70$  in seven scales; the scales “physical functioning” and “role functioning” remain under that criteria (see Table 3).

### Sensitivity

Comparisons of the means showed significant differences for all scales between the two measurement points (see Table 4). The calculations of the

effect sizes demonstrate high effects ( $d \geq 0.8$ ) on the scales “role functioning,” “emotional functioning,” and “global quality of life” and moderate effects ( $0.5 \leq d < 0.8$ ) on the scale “fatigue.” Only small effects were shown on the scales “physical functioning,” “cognitive functioning,” “social functioning,” and “pain”; the scale “nausea/vomiting” did not show any effects. Calculated on the basis of the RCI, the percentage of patients with a higher quality of life varies between 6.1% (“nausea/vomiting”) and 34.2% (“global quality of life”). The correlations of the scales between the two measurement points vary between 0.44 and 0.67.

Because of the described limitations especially regarding the factorial validity, an attempt was made to construct a factorial structure more suitable to the given sample of patients of oncological rehabilitation. For this purpose, a principal compo-

**Table 3.** Scale construction of EORTC QLQ-C30 items; internal consistency, descriptive statistics, and scalability, prior to treatment

Scale (items)	<i>n</i>	Alpha <sup>a</sup>	Floor <sup>b</sup> /ceiling <sup>b</sup> effects		Scale fit
<b>Functioning scales<sup>c</sup></b>					
Physical (1–5)	953	0.57	25%	0%	65%
Role (6, 7)	847	0.36	34%	11%	19%
Cognitive (21–24)	933	0.85	32%	3%	88%
Emotional (20, 25)	953	0.72	5%	5%	100%
Social (26, 27)	945	0.78	18%	12%	100%
Global quality of life (29, 30)	953	0.90	1%	1%	100%
<b>Symptom scales<sup>d</sup></b>					
Fatigue (10, 12, 18)	953	0.86	5%	9%	100%
Nausea/vomiting (14, 15)	963	0.75	74%	2%	100%
Pain (9, 19)	947	0.89	28%	8%	100%

<sup>a</sup>Cronbach’s  $\alpha$  coefficient (internal consistency).

<sup>b</sup>Percentage of the highest and lowest value, respectively.

<sup>c</sup>High scores on the Functioning Scales indicate better health.

<sup>d</sup>High scores on the Symptom Scales indicate worse health.

**Table 4.** Responsiveness indices for scales of the EORTC QLQ-C30

Scale (items)	Beginning of treatment		End of treatment		$r_{t_1, t_2}^b$	Responsiveness indices		
	$M^a$	$SD$	$M^a$	$SD$		$d^c$	RCI <sup>d</sup>	Improved
Functioning scales								
Physical (1–5)	74.2	20.6	80.2	20.6	0.64	0.29	37.4	7.6%
Role (6, 7)	60.1	34.8	88.5	11.7	0.58	0.82	77.2	7.2%
Emotional (20, 25)	50.6	27.2	73.1	22.8	0.50	0.83	29.2	23.0%
Cognitive (21–24)	71.4	28.0	79.2	24.1	0.67	0.28	41.1	5.6%
Social (26, 27)	54.8	32.1	69.3	28.7	0.58	0.45	41.7	14.4%
Global quality of life (29, 30)	50.2	20.6	66.3	19.3	0.50	0.78	18.1	34.2%
Symptom scales								
Fatigue (10, 12, 18)	54.9	28.0	40.3	24.7	0.60	0.52	29.0	25.6%
Nausea/vomiting (14, 15)	8.9	19.9	5.4	15.4	0.44	0.18	27.6	6.1%
Pain (9, 19)	38.0	32.3	28.2	28.3	0.63	0.30	29.7	21.8%

<sup>a</sup>Means prior and after treatment; all mean differences are highly significant ( $p < 0.001$ ).

<sup>b</sup>Correlations of the scales between the two measurement points.

<sup>c</sup>Effect size.

<sup>d</sup>Reliable Change Index.

ment analysis with a subsequent varimax rotation (with replacement of missing values,  $n = 972$ ) was first performed with the number of factors being extracted on the basis of the distribution of eigenvalues indicated in the scree test. For the factor analysis, all items were used, with the exception of the two global items that assess the state of health, because they may not contribute substantially to the construction of individual factors. The principal component analysis results in the presented distribution of eigenvalues (6.9, 2.0, 1.5,

1.4, 1.1, 1.0, 0.9, 0.8, 0.8, 0.7; only the first 10 values mentioned); this justifies a limitation to three factors, explaining 47% of the variance. From each of these three factors, the items with the highest loadings is chosen with the intention of forming scales that have a satisfactory internal consistency of Cronbach's  $\alpha \geq 0.70$ . If the factor loadings are equal, the items with the higher values are selected. According to these selection guidelines, nine items have been chosen (see Table 5) for an abridged scoring procedure.

**Table 5.** Selected items and factor loadings of the remaining nine items of EORTC QLQ-C30 (principal components analysis, varimax rotation,  $N = 972$ )

Original scale	Selected items		Factor 1	Factor 2	Factor 3
Emotional functioning	EF 22	Did you worry?	0.88		
	EF 24	Did you feel depressed?	0.83		
	EF 21	Did you feel tense?	0.83		
Physical functioning	PF 1	Do you have any trouble performing strenuous activities, like carrying a heavy shopping bag or a suitcase?		0.79	
	PF 2	Do you have trouble taking a long walk?		0.78	
Role functioning	RF 6	Were you limited in any way in doing either your work or doing household jobs?		0.65	
	RF 7	Were you limited in pursuing your hobbies or other leisure time activities?		0.50	
Nausea/vomiting	N/V 14	Have you felt nauseated?			0.91
	N/V 15	Have you vomited?			0.87

Variance explained for the factor solution: 65%.

**Table 6.** Scale construction of the remaining scales of EORTC QLQ-C30; Internal consistency, descriptive statistics, and scale fit

Scale (items)	<i>n</i>	Alpha <sup>a</sup>	Floor <sup>b</sup> /ceiling <sup>b</sup> effects		Scale fit
Functioning scales <sup>c</sup>					
Physical and role functioning (1, 2, 6, 7)	838	0.66	18%	9%	83%
Emotional functioning (21, 22, 24)	941	0.84	6%	9%	100%
Global Quality of Life (29, 30)	953	0.90	1%	1%	100%
Symptom scale <sup>d</sup>					
Nausea/vomiting (14, 15)	963	0.75	74%	1%	100%

<sup>a</sup>Cronbach's  $\alpha$  coefficient (internal consistency).

<sup>b</sup>Percentage of the highest and lowest value, respectively.

<sup>c</sup>High scores on the Functioning Scales indicate better health.

<sup>d</sup>High scores on the Symptom Scales indicate worse health.

To check the factor structure of the remaining nine items, a principal component analysis was first calculated using varimax rotation (with replacement of missing values;  $n = 972$ ). Again, both of the global items were excluded from the analysis. A three factorial structure was demonstrated (eigenvalues: 3.1, 1.4, 1.4, 0.9, 0.7, 0.5, 0.4, 0.3, 0.3; only the first 10 results mentioned), explaining 65% of the variance. The first factor covers "emotional functioning," the second factor is made up of items of the initial factors "physical functioning" and "role functioning," the third factor is identical with the factor "nausea/vomiting" of the original version (see Table 5).

Afterward, the postulated scale structure was checked by means of calculating the internal consistencies, the confirmatory factor analysis, and the multitrait approach. Table 6 shows the resulting scale values of the three newly developed scales. In spite of having a low number of items, the internal consistencies are satisfactory or good. In accordance with the prior analysis, the scale "nausea/vomiting" again reveals high floor effects with 74%.

The chosen fit indices within the scope of the conducted confirmatory factor analysis (without re-

placement of missing values,  $n = 755$ ) also show satisfactory results (CFI = 0.93; NNFI = 0.92, RMSEA = 0.07).

Table 7 shows the interscale correlations of the abridged version. Mainly, there are low correlations between most of the scales, except the scale "global quality of life," which highly correlates with the other scales, possibly because of the character of a total score.

The calculation of the mean value of all correlation coefficients (after calculating the Fisher Z-transformed correlation coefficients, without the scale "global quality of life") results in a value of  $r_{\text{mean}} = 0.28$ .

Comparisons of the means showed significant differences for all scales between the two measurement points (see Table 8). The calculations of the effect sizes demonstrate high ( $d \geq 0.8$ ) and moderate ( $0.5 \leq d < 0.8$ ) effects on all scales except the scale "nausea/vomiting," which did not show any effects. Calculated on the basis of the RCI, the percentage of patients with a higher quality of life varies between 6.1% ("nausea/vomiting") and 34.5% ("emotional functioning"). The correlations of the scales between the two measurement points vary between 0.44 and 0.69 (see Table 8).

**Table 7.** Correlations between the remaining scales of EORTC-QLQ-C30

Scale (items)	PRF	EF	N/V
Functioning scales			
Physical and role functioning (PRF, 1, 2, 6, 7)			
Emotional functioning (EF, 21, 22, 24)	0.33		
Symptom scale			
Nausea/vomiting (N/V, 14, 15)	0.21	0.27	0.28

**Table 8.** Responsiveness indices for the remaining scales of EORTC QLQ-C30

Scale (items)	Beginning of treatment		End of treatment		Responsiveness indices			
	<i>M</i> <sup>a</sup>	<i>SD</i>	<i>M</i> <sup>a</sup>	<i>SD</i>	$r_{t_1, t_2}$ <sup>b</sup>	<i>d</i> <sup>c</sup>	RCI <sup>d</sup>	Improved
Functioning scales								
Physical and role functioning (1, 2, 6, 7)	51.2	32.3	86.8	10.8	0.69	1.10	52.2	22.9%
Emotional functioning (21, 22, 24)	47.4	28.4	69.0	24.9	0.52	0.76	31.5	34.5%
Global Quality of Life (29, 30)	50.2	20.6	66.3	19.3	0.50	0.78	18.1	34.2%
Symptom scale								
Nausea/vomiting (14, 15)	8.9	19.9	5.4	15.4	0.44	0.18	27.6	6.1%

<sup>a</sup>Means prior and after treatment; all mean differences are highly significant ( $p < 0.001$ ).

<sup>b</sup>Correlations of the scales between the two measurement points.

<sup>c</sup>Effect size.

<sup>d</sup>Reliable Change Index.

## DISCUSSION

Quality of life represents an essential criterion in evaluating treatment outcome and in quality assurance measures of oncological and palliative care. In this regard, the present study aimed at investigating the psychometric properties of a widely used questionnaire, the EORTC QLQ-C30, to measure changes in the subjectively perceived quality of life within a sample of cancer patients in oncological rehabilitation in Germany with heterogeneous diagnoses.

The results of this study indicate that the EORTC QLQ-C30 shows some potential for improvement, especially with regard to indices of factorial validity. Aaronson's originally published structure with nine factors (Aaronson et al., 1993) could not be confirmed, neither by the principal component analysis, by confirmatory factor analysis, nor by calculating the scale fit, which showed that three of nine scales stay below the required criteria of a scale fit above 90%. In addition, intercorrelations of the scales are partially high, especially between the scales physical functioning/role functioning, physical functioning/fatigue and fatigue/social functioning. However, reliability measurements return satisfactory or even good values on most of the scales; limitations were found on the physical and the role functioning scales. The responsiveness indices suggest that merely the scales "emotional functioning," "fatigue," and "global quality of life" show high sensitivity to change, that is, both high effect sizes and high percentages of improved patients. Descriptive analyses show some restrictions concerning the symptom scale nausea/vomiting; 75% of the patients did not rate any problems with nausea or vomiting. In summary it may be concluded that the original version of the EORTC QLQ-C30 contains

redundancies and reveals poor factorial validity, which is why the instrument appears to be uneconomical for oncological rehabilitation in Germany.

We therefore developed an abridged scoring procedure; the postulated four-factor structure of the extracted 11 items was able to be confirmed within the scope of a confirmatory factor analysis and the multitrait approach. Having a more clearly factorial structure, the reduced scoring procedure covers the impairments in relevant areas of quality of life in patients of oncological rehabilitation and shows satisfactory or even good psychometric characteristics and responsiveness indices. The correlations between the remaining scales are low to moderate. Weak points are, on the one hand, the merely moderate internal consistency of 0.66 in the "physical functioning" scale. This can be explained by, among other things, the dichotomous items used in the scale. However, a first application of the shortened questionnaire, using response options on a 4-point scale, revealed noticeable improvements in the internal consistency of the scale "physical functioning" (Koch, Mehnert, & Petersen, 2002). On the other hand, the scale "nausea/vomiting" has high floor effects and, as a consequence, also shows low sensitivity to change.

This study has a number of potential limitations: First of all, the EORTC Quality of Life Group has developed a modified version of the questionnaire in the meantime, which is why any further developments on version 1 could be considered dispensable. However, because of the still widespread use of version 1 in routine documentations in rehabilitative and palliative care in Germany, the results of this study can be useful when analyzing the collected data. Furthermore, factor analysis results are known to be sample dependent. Thus, it has to be considered that the results obtained in this study



are possibly not to be replicated in further studies. Nevertheless, by using an extensive sample, the generalizability of the research results seems to be acceptable. Moreover, because of the design of the study (pre-post-design without control group), it is not possible to differentiate between an actually low treatment outcome and a lack of sensitivity to change on part of the instrument and it is also difficult to distinguish between variability by chance and treatment success (Schuck, 2000). In addition, it might also be argued that the use of exploratory factor analyses is inappropriate for dichotomous items and therefore the results may be biased (Floyd & Widaman, 1995). However, according to the literature (Kim & Mueller, 1978; Muthen, 1978; Gorsuch, 1983), factor analysis also applies to binary variables or a mixture of binary and continuous variables and the results of the conducted exploratory factor analyses therefore seem to be significant.

In summary, it seems—especially with regard to the growing numbers of questionnaires to be answered by rehabilitation patients in the progress of developing quality assurance programs—that the time gain is considerable in a scoring procedure that has been reduced by 65% of the items. Nevertheless, the abridged scoring procedure is not intended to substitute for any development of a shortened version of the EORTC QLQ-C30 (version 3), which is, as mentioned before, currently being developed by the EORTC Quality of Life Group.

## ACKNOWLEDGMENTS

We are very grateful for the constructive cooperation of all the involved clinics, the helpful support of students Christina Krüger and Miroslaw Witt, and to Dr. Bart for providing the data.

## REFERENCES

- Aaronson, N.K., Ahmedzai, S., Bergman, B., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, 85, 365–376.
- Aaronson, N.K., Bullinger, M., & Ahmedzai, S. (1988). A modular approach to quality of life assessment in cancer clinical trials. *Recent Results in Cancer Research*, 111, 231–249.
- Arbuckle, J.L. & Wothke, W. (1999). *Amos 4.0 Users' Guide*. Chicago: SmallWaters.
- Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In *Testing Structural Equation Models*. Long, J.S. (ed.), pp. 136–162. Newbury Park: Sage.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent und discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.
- Floyd, F.J. & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286–299.
- Gorsuch, R. (1983). *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Hays, R.D. (1991). *Memorandum. Revised Multitrait Analysis Program Software (MAP-R)*. Boston, MA: Health Institute, New England Medical Center.
- Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Jacobson, N.S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- Kaasa, S. & Loge, J.H. (2003). Quality of life in palliative care: Principles and practice. *Palliative Medicine*, 17, 11–20.
- Kim, J.O. & Mueller, C.W. (1978). *Factor Analysis: Statistical Methods and Practical Issues*. Vol. 14. Thousand Oaks, CA: Sage Publications.
- King, M.T. (1996). The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Quality of Life Research*, 5, 555–567.
- Koch, U., Mehnert, A., & Petersen, C. (2002). *Evaluation of the Psycho-Oncological Department of the Herford Clinic*. Hamburg: Institute and Polyclinic for Medical Psychology.
- Krischke, N.R. & Petermann, F. (1995). Quality of life of cancer patients in inpatient rehabilitation. *Praxis der Klinischen Verhaltensmedizin und Rehabilitation*, 32, 291–299.
- Küchler, T., Rappat, S., Holst, K., et al. (1999). Impact of psychotherapeutic support on gastrointestinal cancer patients undergoing surgery: Survival results of a trial. *Journal of Hepatogastroenterology*, 46, 322–335.
- Kyriaki, M., Eleni, T., Efi, P., et al. (2001). The EORTC core quality of life questionnaire (QLQ-C30, version 3.0) in terminally ill cancer patients under palliative care: Validity and reliability in a Hellenic sample. *International Journal of Cancer*, 94, 135–139.
- Liang, M.H. (1995). Evaluating measurement responsiveness. *Journal of Rheumatology*, 22, 1191–1192.
- Liang, M.H. (2000). Longitudinal construct validity: Establishment of clinical meaning in patient evaluative instruments. *Medical Care*, 38(Suppl. 9), 1184–1190.
- Morgan, G. (2000). Assessment of quality of life in palliative care. *International Journal of Palliative Nursing*, 6, 406–410.
- Muthen, B.O. (1978). Contributions to factor analysis of dichotomous items. *Psychometrika*, 43, 511–560.
- Osoba, D., Rodriguez, G., Myles, J., et al. (1998). Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology*, 16, 139–144.
- Osoba, D., Zee, B., Pater, J., et al. (1994). Psychometric properties and responsiveness of the EORTC Quality of Life Questionnaire (QLQ-C30) in patients with breast, ovarian and lung cancer. *Quality of Life Research*, 3, 353–364.

- Paci, E., Miccinesi, G., Toscani, F., et al. (2001). Quality of life assessment and outcome of palliative care. *Journal of Pain and Symptom Management*, 21, 179–188.
- Pfennings, L.E., van der Ploeg, H.M., Cohen, L., et al. (1999). A comparison of responsiveness indices in multiple sclerosis patients. *Quality of Life Research*, 8, 481–489.
- Ringdal, G.I. & Ringdal, K. (1993). Testing the EORTC Quality of Life questionnaire on cancer patients with heterogeneous diagnoses. *Quality of Life Research*, 2, 129–140.
- Ringdal, G.I. & Ringdal, K. (2000). A follow-up study of the quality of life in cancer patients with different prognoses. *Quality of Life Research*, 9, 65–73.
- Schuck, P. (2000). Designs and statistics for assessing change sensitivity of questionnaires in health-related quality of life research. *Zeitschrift für Medizinische Psychologie*, 3, 125–130.
- Schulz, H., Winzer, A., Stump, S., et al. (2001). The effects of psycho-oncological interventions on the quality of life of cancer patients. *Onkologie*, 7, 157–166.
- Stump, S., Koch, U., Baider, L., et al. (1998). The effects of a structured psycho-oncological group program—The German-Israeli intervention study. *Psycho-Oncology*, 7(Suppl. 4), 12–14.
- Weis, J., Moser, M.T., Fachinger, D., et al. (2000). Designs and statistics for assessing change sensitivity of questionnaires in health-related quality of life research. In *Quality of Life Research from a Medical Psychology and Medical Sociology Perspective*. Bullinger, M. (ed.), pp. 43–58. Göttingen: Hogrefe.
- Wiebe, S., Rose, K., Derry, P., et al. (1997). Outcome assessment in epilepsy: Comparative responsiveness of quality of life and psychosocial instruments. *Epilepsia*, 38, 430–438.