PERGAMON

European Journal of Cancer 38 (2002) 339–348

European Journal of Cancer

www.ejconline.com

The comparability of quality of life scores: a multitrait multimethod analysis of the EORTC QLQ-C30, SF-36 and FLIC questionnaires

S. Kuenstner[a,b], C. Langelotz[a], V. Budach[c], K. Possinger[a], B. Krause[b], O. Sezer[a,*]

[a]*Department of Oncology and Hematology, Universitätsklinikum Charité, Humboldt Universität, 10098 Berlin, Germany*
[b]*Department of Psychology, Humboldt Universität, Berlin, Germany*
[c]*Department of Radiotherapy, Humboldt Universität, Berlin, Germany*

Received 6 June 2001; received in revised form 24 September 2001; accepted 12 October 2001

Abstract

This study investigates whether similarly named subscales of three quality of life questionnaires, the European Organization for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30), the Medical Outcome Study Quality of Life Questionnaire Short Form 36 (SF-36) and the Functional Living Index Cancer questionnaire (FLIC) assess similar aspects of the patients' quality of life. A multitrait multimethod analysis on the answers of 234 cancer patients showed that subscale correlations as indicators of convergent validity significantly exceed corresponding correlations of discriminant validity in five of the seven dimensions analysed (physical functioning, emotional functioning, pain, fatigue/vitality and nausea/vomiting). The results of the social functioning and overall health subscales are less clear. Content analysis of the social functioning scales reveals that this domain is differently operationalised in the three questionnaires. Linear regressions of the overall health subscales suggest that patients interpret overall health questions of the three questionnaires differently. The results show that overall health subscales of these three questionnaires cannot be equated, while most specific subscales provide valid results. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: EORTC QLQ-C30; FLIC; Multitrait multimethod analysis; Quality of life; SF-36

1. Introduction

Quality of life has become an important issue in the treatment of cancer. This is reflected by the recommendation of the United States Food and Drug Administration (FDA) to assess the quality of life, in addition to overall survival, as an important measure of treatment efficacy for every new drug therapy developed [1]. Quality of life is now considered a more important aspect than some other traditionally used endpoints like remission rate, disease-free survival and time to progression which do not necessarily reflect a benefit for the patient [2]. Meanwhile, federal institutions, like the US Department of Health and Human Services [3], as well as clinical research groups [4–6] recommend the assessment of quality of life as an important endpoint in clinical trials. In 1996, the American Society of Clinical Oncology stated that quality of life is an important endpoint, secondary in importance only to survival [7].

Although the conceptualisation and measurement of quality of life remains a controversial subject, progress has been made in the development and use of standardised measures, but the comparability across different studies is hampered by more than 70 established generic instruments [8]. Most measures operationalize quality of life according to the World Health Organization (WHO) definition of health [9] with at least a physical, an emotional and a social dimension. Often the instrument also includes items measuring the overall health and the overall quality of life.

Among the few instruments developed with careful attention to the principles of test construction and evaluation are the Medical Outcome Study Short Form 36 (SF-36) [10], the European Organization for Research and Treatment of Cancer Quality of Life core questionnaire (QLQ-C30) [5] and the Functional Living Index Cancer (FLIC) [11]. Despite their different operationalisations of quality of life domains, these ques-

* Corresponding author. Tel.: +49-30-450-513105; fax: +49-30-450-527907.
E-mail address: sezer@charite.de (O. Sezer).

0959-8049/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0959-8049(01)00369-0

tionnaires use several subscales for specific quality of life domains which are labelled similarly (i.e. physical functioning or social functioning in all three questionnaires, and mental health in the SF-36, emotional functioning in the QLQ-C30 and psychological functioning in the FLIC. These similar labels imply related underlying concepts. Therefore, one could expect a convergent validity of similarly named subscales when these questionnaires are compared.

Two comparisons of the SF-36 with QLQ-C30 published in English or German found substantial correlations between related subscales [12,13]. Likewise, two studies investigating the relationship between the QLQ-C30 and the FLIC found high correlations between corresponding scales [14,15]. In most cases, correlations as indicators of convergent validity exceeded indicators of discriminant validity indicating correspondence between homologous scales. We are not aware of any study which compared the SF-36 with the FLIC. In this study, we compare the IQOLA SF-36 German version 1.0 with the German version of the QLQ-C30 version 3 and a German version of the FLIC to investigate whether corresponding subscales of these three instruments measure similar aspects.

2. Patients and methods

2.1. Patients

A convenience sample of patients with different malignant diseases and disease status who attended the Universitätsklinikum Charité, Berlin or the Oskar Ziethen Krankhenhaus, Berlin either as inpatients or outpatients between December 1999 and September 2000 was used. All patients were currently either receiving chemotherapy or radiotherapy. Inclusion criteria were a verified cancer diagnosis, proficiency in German, and an expected survival time of more than 3 months. All patients were informed about the objective of this study before they were asked to fill out the questionnaires.

2.2. Quality of life questionnaires

The EORTC QLQ-C30 is the most commonly used quality of life instrument in European cancer trials. The QLQ-C30 is a cancer-specific, self-administered questionnaire designed for the assessment of general aspects of cancer patients for use in clinical trials [5,16,17]. It contains five functional scales (physical, role, cognitive, emotional and social) and three symptom scales (fatigue, pain and nausea/vomiting), as well as several single-item symptom measures and a global health/quality of life scale. 28 of the 30 questions are constructed as 4-point Likert-scales with verbal anchors for each category, the remaining two questions are 7-point Likert scales with verbal anchors for the end-points. All subscales and individual item scores are transformed to a 0–100 scale with higher values indicating a higher functioning in functional scales and an increased presence of symptoms in symptom scales.

The SF-36 is, with more than 2000 publications, one of the most widely used quality of life instruments worldwide [18–22]. Its 36 questions yield eight multi-item scales (physical functioning, role functioning physical, role functioning emotional, bodily pain, vitality, social functioning, mental health and general health). Six dimensions are formed of Likert-scales with three to six answer categories and verbal anchors for each answer category. Two scales are designed as Gutman scales (with four yes/no items each). The subscores for each subscale are transformed to a 0–100 scale with lower values representing a lower functioning or a lower presence of symptoms.

The FLIC was developed by Schipper and colleagues [11]. It is one of the most frequently used quality of life instruments in the English-speaking world [23–26]. The questionnaire contains 22 items with 7-point Likert/linear analogue scales with verbal anchors on each side. Although the FLIC items were originally summed up into one overall quality of life score, several studies have shown its multi-dimensional nature [11,14,15]. The studies by King and colleagues [14] and Mercier and colleagues [15] were the only studies we found which compared the factor structure of the FLIC with another quality of life questionnaire (in both cases the QLQ-C30). Therefore in this study, we used the factor structure of these two studies in order to compare results across different studies. For this study, we transformed the subscales to 0–100 scales with higher values indicating better functioning or a higher degree of symptoms. Since, to date, there is no official German adaptation of the FLIC, we used the German translation by Schliephake [27].

2.3. Order of questionnaires

The data reported in this paper was part of a larger study. Pretests revealed that many patients needed approximately 1 h to fill out the whole questionnaire package. To avoid losses of motivation and memory effects when answering the questions, the package was administered at two time points. The first questionnaire package included the QLQ-C30 and the SF-36 (always in this order), questions regarding demographic and treatment-related data as well as two different answer formats of a short form quality of life instrument. The second package included the QLQ-C30 and the FLIC (in this sequence), as well as two different answer formats of a short form quality of life questionnaire. Patients were asked to fill out the questionnaires themselves with an interviewer available for questions all the time.

2.4. Quality of life changes between wave 1 and wave 2

The mean time between the first and second wave of administration was 17.0 days. Before the data of both waves could be jointly analysed, potential quality of life changes in between had to be controlled. Wilcoxon signed-rank tests and retest reliabilities for the subscales of the seven QLQ-C30 subscales, which were compared with the other questionnaires, were calculated (Table 1). In no subscale were significant changes of the patient's quality of life found. Because the assumption of a normal distribution is not fulfilled for the QLQ-C30 subscales, Kendall's tau-b correlations as indicators of retest reliability were calculated. The resulting retest reliability coefficients (Table 1) were moderate. Pearson's r would have inflated the coefficients to 0.69–0.86. A study by Hjermstad and colleagues [28] found slightly higher coefficients (Pearson's r 0.63–0.91), but in this study the interval between the two waves of administration was 4 days raising the possibility of memory effects.

Because the retest-reliability coefficients of the QLQ-C30 subscales were in the range of former findings and no significant changes in subscores occurred in between, we decided to analyse the three quality of life questionnaires together. The minor changes between the two waves may result in lower associations between the SF-36 and the FLIC subscores.

2.5. Statistics

Subscales for the QLQ-C30, the SF-36 and the FLIC were computed according to the guidelines of the developers. The scaling assumptions of all three questionnaires were assessed using Cronbach's Alpha reliability coefficients.

To test whether similarly labelled subscales assess the same underlying quality of life domain a multitrait multimethod analysis was used [29]. The multitrait multimethod analysis is a common way to compare subscales (called traits) of different quality of life instruments (called methods). According to this approach hetero-method-mono-trait correlations (gray-shaded cells in Table 2, like the physical functioning scales of the QLQ-C30, the SF-36 and the FLIC) are expected to correlate highly with each other, i.e. they should show convergent validity. On the contrary, hetero-method-hetero-trait correlations (cells with vertical lines in Table 2, like the physical functioning scale of the QLQ-C30 and the social functioning scale of the FLIC) and mono-method-hetero-trait correlations (cells with diagonal lines in Table 2, like the physical and emotional functioning scale of the QLQ-C30) are expected to show low correlations with each other which is called discriminant validity. The limit of previous quality of life studies using this approach has been the lack of clarity under which conditions a researcher judged a correlation matrix to show convergent and discriminant validity.

Table 1
Wilcoxon signed rank test and Kendall's tau-b retest reliabilities of EORTC QLQ-C30 scores

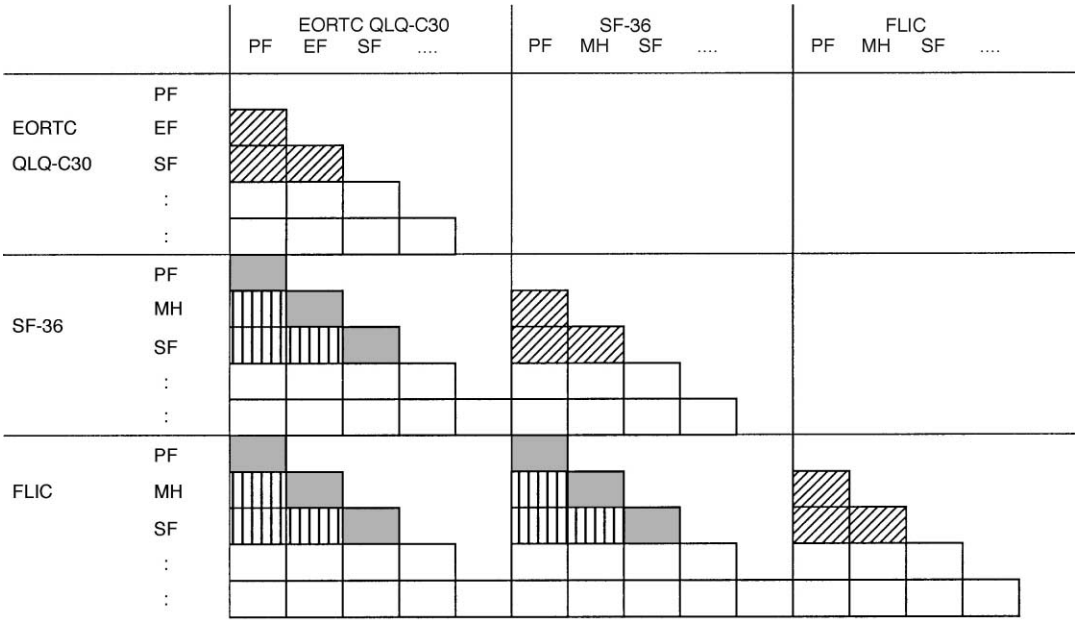
QLQ-C30 dimension	Wilcoxon Z	P value	Retest reliability
Physical functioning	−1.50	0.14	0.72
Emotional functioning	−0.34	0.73	0.65
Social functioning	−0.14	0.89	0.61
Pain	−0.84	0.40	0.65
Fatigue	−1.65	0.10	0.63
Nausea/vomiting	−0.67	0.51	0.67
Quality of life	−1.21	0.23	0.59

EORTC QLQ-C30, European Organization for Research and Treatment of Cancer Quality of Life core questionnaire.

This study makes the decision process explicit by testing the significance of the ratio of correlations of convergent validity (which we will refer to as '*convergent correlations*') that exceed correlations of discriminant validity (which we will refer to as '*discriminant correlations*') in each dimension. For example, the convergent correlation between the physical functioning scales of the QLQ-C30 and the SF-36 was pairwise compared with all discriminant correlations of these two physical functioning scales. This procedure was repeated for the convergent correlations of the QLQ-C30 and the FLIC physical functioning subscales, as well as for the SF-36 and the FLIC subscales. All in all, the three convergent correlations of the physical functioning domain were pairwise compared with 132 discriminant correlations. If the instruments show convergent and discriminant validity, the proportion of convergent correlations exceeding the discriminant correlations should be high.

The ratio of convergent correlations exceeding the discriminant correlations was evaluated using approximative binomial tests. In a binomial test, the probability of getting r observations in one category of a dichotomy and $(n-r)$ observations in the other category, given a sample size n is analysed [30]. The zero hypothesis in our study assumes no differences between convergent and discriminant correlations, i.e. the probability that in a pairwise comparison a convergent correlation exceeds a discriminant correlation is 1/2. The binomial test determines the probability of getting r observations in which the convergent correlations for one quality of life dimension exceed discriminant correlations in this dimension. Since in our study the product of the number of observations and both probabilities $n \times p$ and $n(1-p)$ exceeds 5, the binomial distribution approximates the standard normal distribution, the standard normal distribution was used as an approximation of the binomial test. In case the subscales ana-

Table 2
Scheme of the multitrait multimethod analysis



EORTC QLQ-C30, European Organization for Research and Treatment of Cancer core questionnaire; SF-36, Medical Outcome Study Short Form 36; FLIC, Functional Living Index Cancer.

Methods: the EORTC QLQ-C30, the SF-36, and the FLIC.

Traits: PF (physical functioning), EF (emotional functioning), MH (mental health), MF (mental functioning), and SF (social functioning).

Gray-shaded cells: *hetero-method-mono-trait* correlations, i.e. correlation of similar subscales in different questionnaires, cells with vertical lines: *hetero-method-hetero-trait*, i.e. correlation of different subscales in different questionnaires, cells with diagonal lines: *mono-method-hetero-trait* correlations, i.e. correlation of different subscales in similar questionnaires.

lysed did not show the expected pattern of convergent and discriminant validity, multiple linear regressions were calculated to investigate which subscales of one instrument best explain the variance in the indiscriminant subscales of the other instrument. To test the ability of the instruments in distinguishing between patients with high and low quality of life status, Mann–Whitney U tests between these subgroups of patients were calculated. High quality of life was operationalised as scores above the median in the two QLQ-C30 quality of life scales, the SF-36 overall health scale and the FLIC quality of life scale. Since hardly any of the subscales analysed met the criteria for interval data, the correlations were calculated on the ordinal level as Kendall's tau-b coefficients.

This is the first study that compares the SF-36 and the FLIC. Using the QLQ-C30 as a third questionnaire can give more insight into the likely causes of poor convergent validity. If only two questionnaires are compared, it is unclear whether the low correspondence between related subscales, like in the case of the social functioning scales of the QLQ-C30 and the FACT-G [31], is due to the poor statistical properties of one instrument, whether the two instruments stress different aspects of one domain, or whether a quality of life dimension is more difficult to assess than other dimensions.

3. Results

3.1. Patients

Of the 324 patients addressed, 282 (87%) agreed to participate in this study. Main reasons for non-participation in this study were a disinterest in questionnaire studies, not feeling well and lack of time. The participating patients were asked to fill out a package of questionnaires while they were waiting for their examination, treatment or afterwards to be picked up. 234 (83%) of the 282 patients who filled out the first questionnaire package also took part in the second wave. The patients had a variety of haematological malignancies or solid tumours (Table 3). 217 patients received chemotherapy, 17 patients were treated with radiotherapy. 9 patients (3%) did not want to participate in the second wave, 3 patients (1%) died. The remaining 36 (13%) of the patients failed to fill out their questionnaire because patients were released from the ward or did not have an appointment with their outpatient doctor

Table 3
Characteristics of patients

Characteristics	
Age	
Mean±S.D. (years)	57.6±13.6
Sex	
Female	122
Male	112
Diagnosis	
Breast cancer	86
Gastrointestinal cancer	37
Leukaemia	22
Lung cancer	26
Lymphoma/myeloma	44
Other solid tumours	19

S.D., standard deviation.

within 4 weeks after the first wave. The quality of life scores in wave 1 of those patients who did not take part in the second wave does not differ significantly from those patients who participated in both waves. 47 of the 234 patients (20%) felt unable to fill out the questionnaire alone. In these cases the study was conducted as a structured interview.

3.2. Descriptive data

Median scores and inter-quartile ranges of the transformed subscales of the SF-36, the QLQ-C30, and the FLIC are displayed in Table 4. Most subscales have a median around the centre of the possible answer range indicating no ceiling or bottom effect in this particular patient sample. The social functioning subscale of the FLIC indicates a ceiling effect, the nausea/vomiting subscales of the QLQ-C30 and the FLIC show a bottom effect.

3.3. Multitrait multimethod analysis

The Kendall's tau-b correlation coefficients between the subscales of the three instruments are displayed in Table 5. The diagonal of this correlation matrix contains the Cronbach's Alpha reliability coefficients for the subscales. The Cronbach's Alpha coefficients of all subscales except for the FLIC pain subscale exceeded 0.70, the minimal threshold recommended for multi-item scales. In most pairwise comparisons of quality of life subscale correlations convergent correlations exceeded discriminant correlations (Table 6).

For the quality of life domain physical functioning, the convergent correlations exceeded discriminant correlations in 128 of the possible 132 pairwise comparisons. This significant result ($z=10.71$, $P=0.001$) was unchanged when the analysis was broken down to comparisons between the two questionnaires ($z=6.48$, $P=0.001$ for the comparison of the QLQ-C30 and the SF-36 physical functioning subscale, $z=6.18$, $P=0.001$ for the comparison of the QLQ-C30 and the FLIC subscales, and $z=5.58$, $P=0.001$ for the comparison between the SF-36 and the FLIC subscales). Similar results were found for the emotional functioning and pain subscales. The fatigue/vitality subscale is only part of the QLQ-C30 and the SF-36. Their convergent correlation was higher than 43 of the possible 44 discriminant correlations that involved a fatigue or vitality subscale ($z=6.18$, $P=0.001$). Likewise, the convergent correlation of the nausea/vomiting subscale which is only included in the QLQ-C30 and the FLIC exceeded the corresponding discriminant correlations in all 44 possible comparisons ($z=6.48$, $P=0.001$).

Results were less clear for the quality of life dimensions social functioning and overall health. When analysed over all three questionnaires, convergent

Table 4
Medians and inter-quartile ranges of the EORTC QLQ-C30, the SF-36, and the FLIC

	Wave 1				Wave 2			
	QLQ-C30[a]		SF-36[b]		QLQ-C30[a]		FLIC[a]	
Quality of life dimension	Median	IQR[c]	Median	IQR[c]	Median	IQR[c]	Median	IQRC[c]
Physical functioning	66.7	45.4	50.0	47.2	66.7	46.7	62.5	37.5
Emotional functioning/mental health	66.7	41.7	68.0	32.0	66.7	33.3	66.7	30.0
Social functioning	66.7	50.0	75.0	50.0	66.7	66.7	91.7	33.3
(Bodily) Pain	33.3	50.0	62.0	59.0	33.3	50.0	50.0	41.7
Fatigue (QLQ-C30)/vitality (SF-36)	50.0	44.4	50.0	35.0	44.4	44.4		
Nausea/vomiting	0.0	33.3			0.0	33.3	0.0	25.0
Overall health/quality of life	50.0	33.3	47.0	27.0	50.0	33.3	72.2	33.3

EORTC QLQ-C30, European Organization for Research and Treatment of Cancer core questionnaire; SF-36, Medical Outcome Study Short Form 36; FLIC, Functional Living Index Cancer.

[a] Higher scores indicate a better functioning (quality of life) in functioning scales and a higher degree of symptoms in symptom scales.

[b] Higher scores indicate a better quality of life in all domains.

[c] IQR, inter-quartile range, range 0–100.

Table 5
Multitrait multimethod correlation matrix (Kendall's tau-b correlation coefficients) and Cronbach's Alpha reliability coefficients

		Wave 1													Wave 2												
		EORTC QLQ−C30							SF−36						EORTC QLQ−C30							FLIC					
		PF	EF	SF	PA	FA	NV	QL	PF	MH	SF	BP	VIT	HEA	PF	EF	SF	PA	FA	NV	QL	PF	MF	SF	PA	NAU	HEA
	PF	**0.86**																									
	EF	0.34	**0.79**																								
EORTC	SF	0.34	0.36	**0.73**																							
QLQ−C30	PA	−0.43	−0.35	−0.24	**0.86**																						
(wave 1)	FA	−0.61	−0.45	−0.31	0.45	**0.85**																					
	NV	−0.36	−0.29	−0.20	0.34	0.42	**0.83**																				
	QL	0.47	0.40	0.29	−0.38	−0.50	−0.32	**0.88**																			
	PF	0.68	0.29	0.26	−0.40	−0.50	−0.32	0.47	**0.93**																		
	MH	0.34	0.57	0.34	−0.31	−0.36	−0.27	0.41	0.31	**0.81**																	
SF−36	SF	0.34	0.31	0.54	−0.21	−0.31	−0.22	0.27	0.25	0.39	**0.79**																
(wave 1)	BP	0.38	0.29	0.24	−0.69	−0.40	−0.28	0.36	0.36	0.31	0.25	**0.87**															
	VIT	0.51	0.40	0.29	−0.37	−0.59	−0.35	0.47	0.43	0.45	0.33	0.39	**0.84**														
	HEA	0.40	0.30	0.26	−0.30	−0.44	−0.33	0.43	0.40	0.37	0.25	0.28	0.41	**0.79**													
	PF	0.72	0.27	0.28	−0.39	−0.53	−0.35	0.44	0.63	0.31	0.33	0.42	0.47	0.38	**0.88**												
	EF	0.30	0.65	0.30	−0.31	−0.37	−0.30	0.40	0.29	0.52	0.29	0.31	0.36	0.29	0.36	**0.81**											
EORTC	SF	0.31	0.37	0.61	−0.24	−0.29	−0.18	0.22	0.23	0.37	0.53	0.28	0.27	0.18	0.36	0.38	**0.85**										
QLQ−C30	PA	−0.35	−0.31	−0.18	0.65	0.36	0.30	−0.36	−0.36	−0.28	−0.22	−0.67	−0.31	−0.26	−0.47	−0.38	−0.28	**0.89**									
(wave 2)	FA	−0.53	−0.39	−0.24	0.36	0.63	0.38	−0.44	−0.46	−0.34	−0.31	−0.39	−0.56	−0.40	−0.62	−0.50	−0.35	0.45	**0.88**								
	NV	−0.32	−0.27	−0.15	0.27	0.36	0.67	−0.34	−0.29	−0.26	−0.21	−0.27	−0.32	−0.26	−0.36	−0.36	−0.19	0.29	0.41	**0.77**							
	QL	0.41	0.35	0.26	−0.33	−0.40	−0.20	0.58	0.36	0.38	0.28	0.36	0.42	0.34	0.42	0.38	0.30	−0.36	−0.43	−0.26	**0.90**						
	PF	0.56	0.25	0.28	−0.24	−0.44	−0.29	0.36	0.48	0.33	0.32	0.31	0.44	0.30	0.60	0.35	0.33	−0.31	−0.49	−0.32	0.35	**0.81**					
	PF	0.27	0.44	0.27	−0.22	−0.24	−0.22	0.31	0.26	0.54	0.27	0.25	0.30	0.25	0.28	0.51	0.37	−0.24	−0.27	−0.24	0.29	0.34	**0.79**				
FLIC	SF	0.10	0.08	0.14	−0.06	−0.13	−0.02	0.09	0.06	0.14	0.18	0.09	0.17	0.12	0.16	0.14	0.21	−0.13	−0.16	−0.03	0.12	0.16	0.14	**0.71**			
(wave 2)	PA	−0.41	−0.40	−0.32	0.38	0.40	0.33	−0.38	−0.34	−0.40	−0.30	−0.42	−0.41	−0.28	−0.46	−0.41	−0.38	0.44	0.43	0.30	−0.42	−0.45	−0.33	−0.20	**0.59**		
	NAU	−0.26	−0.22	−0.12	0.19	0.29	0.53	−0.27	−0.21	−0.29	−0.16	−0.23	−0.26	−0.22	−0.28	−0.33	−0.16	0.21	0.36	0.66	−0.25	−0.32	−0.22	−0.07	0.35	**0.88**	
	HEA	0.37	0.25	0.19	−0.19	−0.40	−0.24	0.36	0.33	0.36	0.25	0.26	0.44	0.26	0.40	0.37	0.27	−0.28	−0.42	−0.30	0.40	0.49	0.35	0.18	−0.40	−0.30	**0.90**

EORTC QLQ-C30, European Organization for Research and Treatment of Cancer core questionnaire; SF-36, Medical Outcome Study Short Form 36; FLIC, Functional Living Index Cancer. EORTC QLQ-C30: PF, physical functioning; EF-emotional functioning; SF, social functioning; PA, pain; FA, fatigue; NV, nausea/vomiting, QL, overall health/quality of life.

SF-36: PF, physical functioning; MH, mental health; SF, social functioning; BP, bodily pain; VIT, vitality; HEA, overall health.

FLIC: PF, physical functioning; MF, mental functioning; SF, social functioning; PA, pain; NAU, nausea; HEA, overall health.

High correlations are expected between pairs of subscales measuring the same quality of life dimension (outlined cells) to indicate convergent validity. The remaining off-diagonal cells are expected to show low correlations to indicate discriminant validity. The diagonal shows the Cronbach's Alpha internal consistency coefficients.

Table 6
Number of comparisons where subscale correlations of convergent validity exceed subscale correlations of discriminant validity

Quality of life dimension	Subscales of the instruments			Number			z	P value
Physical functioning	QLQ-C30	with	SF-36	44	of	44	6.48	0.001
	QLQ-C30	with	FLIC	43	of	44	6.18	0.001
	SF-36	with	FLIC	41	of	44	5.58	0.001
	overall			128	of	132	10.71	0.001
Emotional functioning/ mental health	QLQ-C30	with	SF-36	44	of	44	6.48	0.001
	QLQ-C30	with	FLIC	44	of	44	6.48	0.001
	SF-36	with	FLIC	44	of	44	6.48	0.001
	overall			132	of	132	11.40	0.001
Social functioning	QLQ-C30	with	SF-36	44	of	44	6.48	0.001
	QLQ-C30	with	FLIC	27	of	44	1.36	0.174
	SF-36	with	FLIC	21	of	44	−0.45	0.653
	overall			92	of	132	4.44	0.001
Pain	QLQ-C30	with	SF-36	44	of	44	6.48	0.001
	QLQ-C30	with	FLIC	40	of	44	5.28	0.001
	SF-36	with	FLIC	39	of	44	4.97	0.001
	overall			123	of	132	9.84	0.001
Fatigue/vitality	QLQ-C30	with	SF-36	43	of	44	6.18	0.001
Nausea/vomiting	QLQ-C30	with	FLIC	44	of	44	6.48	0.001
Overall health/quality of life	QLQ-C30	with	SF-36	37	of	44	4.37	0.001
	QLQ-C30	with	FLIC	32	of	44	2.86	0.004
	SF-36	with	FLIC	12	of	44	−3.17	0.002
	overall			81	of	132	2.52	0.012

correlations in the domain social functioning significantly exceeded corresponding discriminant correlations ($z = 4.44$, $P = 0.001$). Yet when broken down to pairwise comparisons of two questionnaires, only the association between the QLQ-C30 and the SF-36 social functioning subscales remained superior to discriminant correlations ($z = 6.48$, $P = 0.001$). Neither the correlation between the QLQ-C30 and the FLIC social functioning subscale nor the correlation between the SF-36 and the FLIC social functioning subscale significantly exceeded the discriminant correlations of the social functioning subscales with unrelated subscales.

Likewise, the convergent correlations between overall health subscales significantly exceeded discriminant correlations of the overall health subscales with unrelated subscales ($z = 2.52$, $P = 0.01$). But when two questionnaires were compared only the association between the QLQ-C30 quality of life subscale and the overall health subscales of the SF-36 and the FLIC were significantly larger than their discriminant correlations. On the contrary, the correlation between the SF-36 and the FLIC overall health subscale was significantly lower than the associated discriminant correlations.

3.4. Linear regressions for indiscriminant subscales

To investigate which other quality of life subscales explain more variance in the social functioning and overall health scales than the corresponding subscales of the other two questionnaires, multiple linear regressions with the standardised rank transformed subscales of the other instruments as predictor variables were calculated. When the QLQ-C30 subscales were used as predictor variables of the FLIC social functioning subscale, the QLQ-C30 social functioning subscale was the only significant predictor. But it accounted for only 7.0% of the variance. Likewise, the SF-36 social functioning scale was the only significant SF-36 predictor variable of the FLIC social functioning scale ($R^2 = 0.063$). Two FLIC subscales were significant predictors of the QLQ-C30 social functioning subscale: the FLIC pain (24.0%) and mental functioning subscale (8.3%), yielding a total of 32.3% explained variance. Three FLIC subscales explained 22.1% of the variance in the SF-36 social functioning subscale, physical functioning (17.0%), mental functioning (3.4%), and pain (an additional 1.7%). Neither for the QLQ-C30 nor for the SF-36 social functioning subscale was the FLIC social functioning subscale a significant predictor.

The SF-36 overall health subscale was best predicted by the FLIC physical subscale ($R^2 = 0.175$). The FLIC pain and mental health subscales added 3.3 and 2.0%, respectively. 33.8% of the FLIC overall health score was predicted by the SF-36 vitality subscore. The SF-36 overall health subscore and the SF-36 physical functioning subscore explained an additional 3.1 and 2.5%, respectively.

3.5. Ability of the instruments to distinguish between patients with high and low quality of life status

The high quality of life group consisted of 44 patients that scored above the median in the two QLQ-C30 quality of life subscales, the SF-36 overall health scale, and the FLIC quality of life scale. The low quality of life group was formed of the 48 patients that scored below the median in all four indicators of overall quality of life. The discriminant ability of the three questionnaires in distinguishing between patients with varying quality of life status was tested on these two subgroups. Mann–Whitney U tests showed significant differences in the rank sums of these two subgroups (Table 7). In all quality of life domains, patients with higher quality of life status reported a better functioning or less symptoms than patients with a low quality of life status.

4. Discussion

This study investigated whether seven similarly named subscales of the SF-36, the QLQ-C30 and the FLIC measure the same underlying trait. When analysing the multitrait multimethod correlation matrix the decision process was made explicit by testing the significance of the ratio of convergent correlations exceeding discriminant correlations in each dimension. It is the first study that compared three standardised multidimensional quality of life questionnaires and the first study comparing the SF-36 and the FLIC.

Five of the seven quality of life dimensions showed convergent validity. Like in previous studies, the physical dimensions (physical functioning, pain, fatigue, and nausea/vomiting) and the emotional functioning subscales correlate moderately to highly with each other [12–15,31,32]. More importantly, in these dimensions convergent correlations between similarly labelled subscales correlated higher with each other than with unrelated subscales. In addition, the significant differences between patients with high and low quality of life status in all quality of life domains implies that all three instruments are able to discriminate between patients with varying quality of life status.

The low internal consistency of the FLIC pain subscale is most probably due to an unprecise translation of item 20. While the English version asks for pain and discomfort the German translation [27] is more general asking for painful/sorrowful discomfort. Despite its low internal consistency, the FLIC pain subscale showed good convergent and discriminant validity.

The poor convergent correlation of the social subscales is in line with previous findings that found lower correlations between social subscales compared with physical domains [12,32]. A multiple linear regression model yielded no other significant predictors of the social subscales. This finding leads to the conclusion that the low convergent validity of the QLQ-C30, SF-36 and FLIC social subscales is less due to an overlap with presumably unrelated traits, but result from a lack of overlap between these two social functioning scales. A possible explanation would be the way how social functioning is operationalised. While the QLQ-C30 and the SF-36 ask for disruptions in relations with friends and relatives due to health problems, the FLIC asks the patient to rate his willingness to see and spend time with relatives and friends. Apparently these operationalisations of social functioning activate different knowledge structures which yield different results and possibly diverging conclusions.

Overall health ratings are included in quality of life measures to allow for weightings of different health domains and to measure health aspects that are not covered by the specific questions. Generally overall health ratings correlate highly with each other [32]. Despite their general character, those dimensions correlated only moderately with each other in our study. Especially for the convergent correlation between the SF-36 and the FLIC, the overall health ratings was poor and significantly lower than correlations of these subscales with more specific quality of life aspects. The

Table 7
Discriminant ability of the questionnaires in distinguishing between patients with a high and low quality of life status[a]

Quality of life dimensions	Wave 1				Wave 2			
	QLQ-C30		SF-36		QLQ-C30		FLIC	
	U	*P* value	U	*P* value	U	*P* value	U	*P* value
Physical functioning	157.0	0.001	195.5	0.001	129.5	0.001	159.5	0.001
Emotional functioning/mental health	346.5	0.001	185.5	0.001	220.5	0.001	249.5	0.001
Social functioning	527.5	0.001	473.0	0.001	475.5	0.001	564.5	0.001
Pain	457.5	0.001	402.5	0.001	400.5	0.001	292.0	0.001
Fatigue/vitality	118.5	0.001	91.5	0.001	83.0	0.001		
Nausea/vomiting	497.5	0.001			455.5	0.001	470.5	0.001

[a] Mann–Whitney U tests for differences in subgroups were significant with all of the three quality of life instruments used in this study.

ordinal multiple linear regression revealed that the SF-36 overall health subscale is best predicted by the physical subscale of the FLIC. Conversely, the FLIC overall health subscore is best explained by the SF-36 vitality subscore. The SF-36 overall health questions inquire about the patient's health directly or as comparisons with other persons. Apparently patients associate these questions more with physical aspects. Conversely, the FLIC overall health questions ask about the patients well-being and feelings, which apparently is more associated with subjective feelings of vitality/fatigue. The QLQ-C30 tries to find a compromise by asking one question (item 29) about the patient's health and a second question (item 30) concerning the patient's overall quality of life. Presumably this hybrid approach leads to higher convergent validity when compared with the SF-36 and the FLIC overall health subscales.

Although the use of quality of life assessment has become common in clinical trials, many studies do not use standardised multidimensional quality of life measures or use only overall health/quality of life ratings when they report about the patient's well-being. This study shows that careful interpretation is recommended when measures of so-called 'overall health' are compared. Although all three questionnaires in our study show a high internal consistency in this scale they focus on different aspects of the overall well-being. Comparisons of overall health scores across studies that use different quality of life instruments may lead to the wrong conclusions, whereas most subscales of the three measures analysed showed a high degree of convergent and discriminant validity in the seven quality of life dimensions studied.

Acknowledgements

This study was supported by a research grant from the Hector Stiftung, Weinheim, Germany.

References

1. Johnson JR, Temple R. Food and Drug Administration requirements for approval of new anticancer drugs. *Cancer Treat Rep* 1985, **69**, 1155–1159.
2. Beitz J, Gnecco C, Justice R. Quality-of-life end points in cancer clinical trials: the U.S. Food and Drug Administration perspective. *J Natl Cancer Inst Monogr* 1996, **20**, 7–9.
3. Division of Cancer Prevention and Control. *Quality of life assessment in cancer clinical trials. Report of the workshop on quality of life research in cancer clinical trials*. Bethesda, MD, National Cancer Institute, 1990.
4. Osoba D, Aaronson NK, Till JE. A practical guide for selecting quality of life measures in clinical trials and practice. In Osoba D, ed. *Effect of Cancer on Quality of Life*. Boston, CRC Press, 1992, pp 89–104.
5. Aaronson NK, Ahmedzai S, Bergmann B, *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**, 365–376.
6. Moinpour CM. Measuring quality of life: an emerging science. *Semin Oncol* 1994, **21**, 48–60.
7. Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. American Society of Clinical Oncology. *J Clin Oncol* 1996, **14**, 671–679.
8. Sloan JA, Loprinzi CL, Kuross SA, *et al*. Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. *J Clin Oncol* 1998, **16**, 3662–3673.
9. World Health Organization. *The constitution of the World Health Organization*. Geneva, WHO Basic Documents, 1948.
10. Brazier JE, Harper R, Jones NM, *et al*. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *Br Med J* 1992, **305**, 160–164.
11. Schipper H, Clinch J, McMurray A, Levitt M. Measuring the quality of life of cancer patients: the Functional Living Index-Cancer: development and validation. *J Clin Oncol* 1984, **2**, 472–483.
12. Apolone G, Filiberti A, Cifani S, Ruggiata R, Mosconi P. Evaluation of the EORTC QLQ-C30 questionnaire: a comparison with SF-36 Health Survey in a cohort of Italian long-survival cancer patients. *Ann Oncol* 1998, **9**, 549–557.
13. Krischke NR. *Lebensqualität und Krebs*. München, Quintessenz, 1996.
14. King MT, Dobson AJ, Harnett PR. A comparison of two quality-of-life questionnaires for cancer clinical trials: the functional living index-cancer (FLIC) and the quality of life questionnaire core module (QLQ-C30). *J Clin Epidemiol* 1996, **49**, 21–29.
15. Mercier M, Bonneterre J, Schraub S, Lecomte S, el Hasnaoui A. The development of a French version of a questionnaire on the quality of life in cancerology (Functional Living Index-Cancer: FLIC). *Bull Cancer* 1998, **85**, 180–186.
16. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res* 1996, **5**, 555–567.
17. Osoba D, Zee B, Pater J, Warr D, Kaizer L, Latreille J. Psychometric properties and responsiveness of the EORTC quality of Life Questionnaire (QLQ-C30) in patients with breast, ovarian and lung cancer. *Qual Life Res* 1994, **3**, 353–364.
18. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992, **30**, 473–483.
19. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994, **32**, 40–66.
20. Ware JE Jr, Kosinski M, Gandek B, *et al*. The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998, **51**, 1159–1165.
21. Stewart AL, Hays RD, Ware JE Jr. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* 1988, **26**, 724–735.
22. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ* 1993, **2**, 217–227.
23. Ganz PA, Lee JJ, Siau J. Quality of life assessment. An independent prognostic variable for survival in lung cancer. *Cancer* 1991, **67**, 3131–3135.
24. Ganz PA, Haskell CM, Figlin RA, La-Soto N, Siau J. Estimating the quality of life in a clinical trial of patients with metastatic lung cancer using the Karnofsky performance status and the Functional Living Index-Cancer. *Cancer* 1988, **61**, 849–856.
25. Cassileth BR, Lusk EJ, Guerry D, *et al*. Survival and quality of life among patients receiving unproven as compared with conventional cancer therapy. *N Engl J Med* 1991, **324**, 1180–1185.
26. Morrow GR, Lindke J, Black P. Measurement of quality of life

in patients: psychometric analyses of the Functional Living Index-Cancer (FLIC). *Qual Life Res* 1992, **1**, 287–296.
27. Schliephake H, Neukam FW, Schmelzeisen R, Varoga B, Schneller H. Long-term quality of life after ablative intraoral tumour surgery. *J Craniomaxillofac Surg* 1995, **23**, 243–249.
28. Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S. Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire. *J Clin Oncol* 1995, **13**, 1249–1254.
29. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959, **56**, 81–105.
30. Siegal S, Castellan NJ. *Nonparametric Statistics for the Behavioral Sciences*, 2nd edn. New York, McGraw-Hill, 1998.
31. Kemmler G, Holzner B, Kopp M, *et al*. Comparison of two quality-of-life instruments for cancer patients: the functional assessment of cancer therapy-general and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30. *J Clin Oncol* 1999, **17**, 2932–2940.
32. Kopp M, Schweigkofer H, Holzner B, *et al*. EORTC QLQ-C30 and FACT-BMT for the measurement of quality of life in bone marrow transplant recipients: a comparison. *Eur J Haematol* 2000, **65**, 97–103.