Eur J Cancer Clin Oncol, Vol. 24, No. 4, pp. 691–701, 1988.
Printed in Great Britain

0277-5379/88$3.00+0.00

Validation of a Quality of Life Questionnaire for Use in Clinical Trials for Treatment of Patients with Inoperable Lung Cancer*

STEIN KAASA,† ARNE MASTEKAASA,‡ INGER STOKKE† and SIRI NAESS§

Department of Medical Oncology and Radiotherapy, †The Norwegian Radium Hospital, Oslo, ‡Institute for Social Research, Oslo and §Institute of Applied Social Research, Oslo, Norway

Abstract—*In order to assess the quality of life of patients with inoperable lung cancer, a questionnaire for patient self-administration with 29 variables was designed including the following subjects: psychosocial well-being, medical side-effects, activities of daily living and physical performance. The questionnaire was validated by using a semistructured interview. Thirty-one patients treated for inoperable lung cancer with either radiotherapy (42%), chemotherapy (42%) or a combination of these two treatment modalities (16%) were included in the study. The multitrait–multimethod matrix (MTMM) was used in the analysis of the construct validity. The present validation study showed a high degree of validity for the majority of the items studied. Some of the items showed a low degree of validity, and are subject to future analysis in large scale studies.*

INTRODUCTION

IN ONCOLOGY the effects of different treatments on cancer are evaluated in clinical trials. Response rate and survival are the usual end-points. For the large group of patients with non-curable disease, who have only marginal benefit from chemo- and/or radiotherapy [1, 2], there is an increasing need for methods to measure quality of life.

In 1949 Karnofsky [3, 4] developed the so-called Karnofsky performance status. In many studies this performance status has been used as the only indicator of quality of life. However, its validity and reliability are still debated [5, 6]. During the last decade increased effort has been made to determine more directly a patient's quality of life. However, a generally accepted method to measure quality of life is not yet available [7].

Lung cancer represents one of the most frequent neoplastic diseases [8], and is often incurable at the time of presentation. Combination chemotherapy containing cisplatin has shown response rates from 20 to 50%; however, the influence on survival is marginal [9]. In clinical trials a somewhat higher response rate has been obtained by combined radiotherapy and chemotherapy, while the effect of radiotherapy on survival is doubtful [10]. Both radiotherapy and combination chemotherapy have considerable side-effects. In lung cancer patients it seems important to know how various treatments influence the person's quality of life in order to make rational decisions regarding treatment choice.

The aim of the present study is to validate a method for quality of life measurements that can be used in clinical trials evaluating different treatments of patients with inoperable lung cancer. Validity can be defined as the extent to which the result of a measurement method measures what it is meant to measure. (Reliability concerns the extent to which a test yields the same results on repeated trials.) The patients' responses to a standardized self-administered questionnaire are compared to the results of a semistructured interview with the same patients.

MATERIALS AND METHODS

Patient characteristics

Thirty-one patients with inoperable lung cancer, nine women and 22 men, hospitalized at The Norwegian Radium Hospital, were included in the study. The mean age was 61 years (47–77 years), 13 (42%) were treated with radiotherapy to the hilar region and mediastinum (2.8 Gy × 15/21 days) or 3 Gy × 10/14 days). Thirteen (42%)

Accepted 4 November 1987.
*Supported by The Norwegian Cancer Society, Oslo.
Reprint requests to: Stein Kaasa, The Norwegian Radium Hospital, Department of Medical Oncology and Radiotherapy, Montebello, 0310 Oslo 3, Norway.

DURING THE LAST FORTNIGHT WOULD YOU SAY THAT YOU .....

	Never	Occasionally	Some of the time	Often	All the time
S 1 Have felt strong and energetic?					
S 2 Have felt lonely?					
S 3 Have felt close to/intimate with another person?					
S 4 Have felt tired and rundown?					
S 5 Have felt satisfied with yourself?					
S 6 Have felt that life is worth living?					
S 7 Have been in a good mood?					
S 8 Have lacked confidence in yourself?					
S 9 Have felt depressed?					
S 10 Have felt that life is meaningless?					

Fig. 1. Psychosocial well-being index. Page one of the questionnaire: it was presented to the patients in Norwegian, translation into English is made for publication.

received chemotherapy with either a combination of cisplatin (70–100 mg/m^2) and etoposide (100 mg/m^2) or doxorubicin (50 mg/m^2), cyclophosphamide (1000 mg/m^2) and vincristine (2 mg). Five patients (16%) received a combined treatment, i.e. radio- and chemotherapy. Twenty-seven patients had non-small-cell lung cancer while the remaining four patients had small-cell lung cancer.

Design of the questionnaire

A questionnaire was designed with 29 variables, covering the following subjects:

1. Psychosocial well-being (Figs. 1 and 2).
2. Treatment and disease related symptoms (Fig. 3).
3. Activities of daily living and physical performance (Fig. 4).

The selection of psychosocial items was mainly based on previous quality of life studies in general population samples [11, 12]. The psychosocial well-being scale consists of ten items, five positive and five negative items making up five pairs of items each covering a particular psychosocial dimension. This was done in order to correct for the yea-saying/nay-saying bias. A 'yea-sayer' is a person who seems to agree with most items irrespective of content, whereas 'nay-saying' is the corresponding tendency to disagree with most items in a scale.

Additionally, two general quality of life questions were included in the questionnaire. With regard to the disease and treatment-related symptoms the selection of items was based on the clinical experience and the literature on clinical trials in patients with lung cancer [13–15]. The physical performance scale and activity of daily living scale was influenced by the EORTC lung cancer questionnaire which has been found valid in a previous study [16].

The interview

The questionnaire results (Method I) were obtained 1 h before the interview. They were compared to the ratings based on a 45–60 min semi-structured interview with each patient. The interviews were carrried out by a trained interviewer. These interviews were designed to provide more detailed information about the subjects covered in the questionnaire. The interviews were taped in order to be available for later validation procedures. After finishing the interview, the interviewer filled out an identical questionnaire (Method II). Experienced oncologists listened carefully to the taped interviews and completed an identical questionnaire (Method III). This gives three different methods (I, II and III) for analysing the validity.

A multitrait–multimethod matrix was constructed [17] according to Campbell and Fiske, in order to compare the different methods (Fig. 5).

Taking all things together, how happy would you say you have been the last fortnight?

1. Very happy _____
2. Happy _____
3. Somewhat happy _____
4. Mixed _____
5. Somewhat unhappy _____
6. Unhappy _____
7. Very unhappy _____

Thinking about how life has been the last fortnight. Are you generally satisfied or dissatisfied?

1. Very satisfied _____
2. Satisfied _____
3. Somewhat satisfied _____
4. Mixed _____
5. Somewhat dissatisfied _____
6. Dissatisfied _____
7. Very dissatisfied _____

Fig. 2. Psychosocial well-being, global questions. Page two of the questionnaire.

Did you have any of the following symptoms during the last fortnight?

	Never	Little	Quite a bit	Very much
M 1 Did you have nausea?				
M 2 Did you vomit?				
M 3 Did you have hairloss?				
M 4 Did you have trouble swollowing?				
M 5 Did you have sore throat?				
M 6 Did you have sore and red skin?				
M 7 Were you tired?				
M 8 Did you lack appetite?				
M 9 Did you have trouble sleeping?				
M 10 Did you have pain?				

Fig. 3. Treatment- and disease-related symptoms. Page three of the questionnaire.

	Yes	No
A 1 Do you have to stay indoors most of the day?	___	___
A 2 Are you in bed most of the time?	___	___
A 3 Do you have any trouble climbing one flight of stairs?	___	___
A 4 Do you need help to bath, dress or using the toilet?	___	___

Are you short of breath when you .....

	Not at all	Little	Quite a bit	Very much
F 1 Are at rest?				
F 2 Walk?				
F 3 Climb stairs?				

Fig. 4. Activities of daily living and physical performance. Page four of the questionnaire.

Statistical procedures

Campbell and Fiske developed four criteria to assess the magnitude of relationships between traits and indicators (multitrait–multimethod matrix, Fig. 5).

Criterion 1. The correlations between the same trait measured by different methods (validity diagonal values) should be statistically significant and sufficiently large to encourage further examination of validity.

Criterion 2. A validity diagonal value of an indicator should be higher than the correlations between that indicator and any other indicator having neither trait nor method in common. In more technical terms the validity diagnonal value should be higher than the values of the same column and row in the so-called heterotrait–heteromethod triangles.

Criterion 3. The correlation between two indicators of the same trait measured by different methods should be greater than correlations between different traits measured by the same method. That is, the validity diagnonal values should be higher than the heterotrait–monomethod row or column for the given trait.

Criterion 4. A similar pattern of trait intercorrelations should be shown in both monomethod and heteromethod blocks.

In order to facilitate comparison, factor analysis [18–20] of the intercorrelations among the indicators may have been a favourable method. However, in the present study with only 31 patients, the number of patients may be regarded as too small to carry out factor analyses.

More simple and less rigorous methods may therefore be preferable. To the extent that the pattern of trait intercorrelations is similar irrespective of which method is used, the direction of the intercorrelations among the items should be consistent. In order to sum up the similarities for the psychosocial well-being questions one counted how often each trait intercorrelation had a consistent positive or negative sign across the blocks. With nine blocks, this gives a maximum of nine and a minimum of five similarities.

Reliability may be defined as the degree of random error in a measure. Generally it may be checked by repeated administrations of the same tests (test–retest reliability), or by analyzing the internal consistency in a set of related measures. Validity, as stated in the Introduction, may be defined as the extent to which a result of a measurement method measures what it is meant to measure. It follows that both the random and systematic error addresses

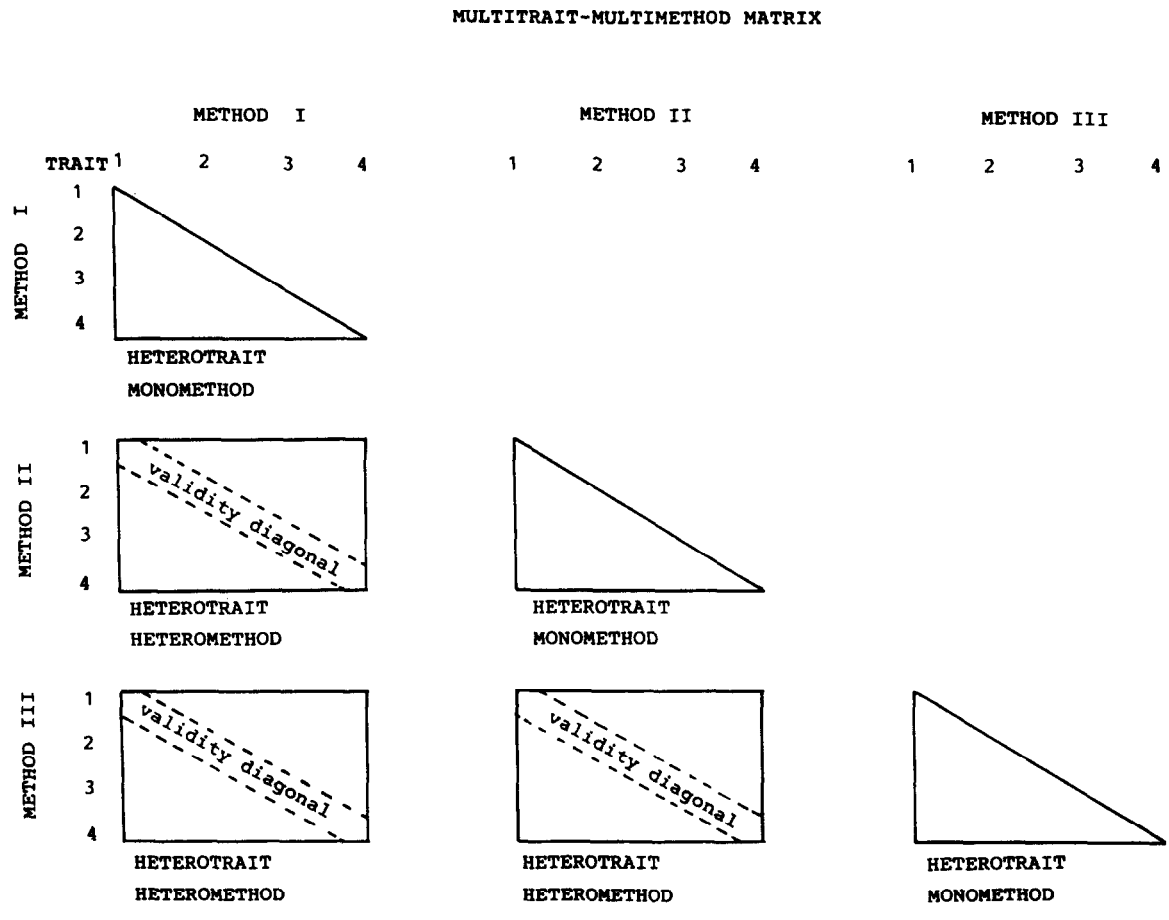

Fig. 5. Basic nomenclature of the MTMM matrix: Three triangles called the heterotrait–monomethod blocks are made, that is correlations between different traits measured by the same method. Three rectangles called the heterotrait–heteromethod blocks are made. They represent the correlations between different traits measured by different methods. The validity diagonal values are correlations between the same trait measured by different methods.

validity and should be measured. In the present study, however, we do not have repeated tests for each patient. It may not be possible to do test–retest reliability for such a fluctuating variable as psychosocial well-being or treatment-related side-effects with a test–retest interval of 1–2 weeks. Furthermore, the internal consistency of the different questions should be checked in a large scale study after the discriminant and convergent validity have been analyzed as in the present study.

The Peerson product moment correlation (r) [21] is used. This is a measure of the linear relationship between two variables. If r is close to zero there is a low or no relationship between the variables. If the value r is close to $+1$ or -1 there is a strong relationship. The object of correlation analysis is to determine the extent to which variation in one variable is associated with variation in the other.

RESULTS

Psychosocial well-being

In Table 1 the design of the multitrait–multimethod is shown. Interpretation of the matrix is presented in Tables 2–5. Table 2 presents the validity diagonal values (correlations between the same trait measured by different methods) which turn out to be rather high for the majority of the items. However, three of the items do not show significant correlations at the 5% level.

For item S5 'pleased with yourself' the patient/interviewer correlation is 0.23 and not significant. As far as S6, 'feeling that life is worth living' is concerned the correlations between self-rating and interview rating and between self-rating and observer-rating are both negative, whereas the correlation between interviewer and observer is positive and significant ($r = 0.60$).

Item S8 'lacked confidence in yourself' show a significant correlation (0.38) between the patient and the interviewer, but the patient–observer correlation is low (0.15) and not statistically significant.

The second validity criterion specifies that the validity diagonal values should be higher than the corresponding values in the heterotrait–heteromethod row and column of the multitrait–multimethod matrix. These comparisons are shown in Table 3.

Generally the validity requirement is met in the comparisons, with items S5, S6 and S8 again being

Table 1. *Multitrait–multimethod psychosocial well-being index*

Traits		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S1	S2	S3	S4
Method I	S1	1.00													
	S2	−0.23	1.00												
	S3	0.26	−0.40	1.00											
	S4	0.52	−0.26	0.08	1.00										
	S5	−0.66	0.20	−0.12	−0.45	1.00									
	S6	0.18	−0.60	0.39	0.05	−0.11	1.00								
	S7	0.41	−0.18	0.06	0.51	−0.43	0.05	1.00							
	S8	−0.47	0.18	−0.45	−0.36	0.28	−0.30	−0.35	1.00						
	S9	−0.33	0.32	−0.12	−0.33	0.25	−0.18	−0.54	0.15	1.00					
	S10	−0.03	0.03	−0.17	−0.22	−0.34	−0.23	−0.45	0.05	0.44	1.00				
Method II	S1	0.28	−0.06	0.07	−0.72	0.43	0.13	0.47	−0.33	−0.18	−0.30	1.00			
	S2	−0.36	0.43	−0.33	0.43	−0.29	−0.33	−0.61	0.19	0.13	0.28	−0.36	1.00		
	S3	0.22	−0.01	0.47	−0.02	0.14	0.33	0.30	−0.21	−0.18	−0.20	0.05	−0.28	1.00	
	S5	0.47	0.01	0.02	−0.32	0.23	0.04	0.33	−0.12	−0.13	0.03	0.50	−0.21	−0.23	−0.38
	S6	0.16	0.26	−0.20	−0.22	−0.02	−0.12	0.36	−0.43	−0.02	0.00	0.18	−0.06	−0.12	−0.22
	S7	0.44	−0.16	0.17	−0.30	0.48	0.17	0.73	−0.60	−0.33	−0.19	0.49	−0.37	0.31	−0.48
	S8	−0.29	0.00	0.19	0.27	−0.28	−0.09	−0.46	0.38	0.26	0.24	−0.28	0.10	0.08	0.30
	S9	−0.24	0.10	−0.10	0.45	−0.26	−0.12	−0.51	0.35	0.44	0.40	−0.30	0.25	0.07	0.25
	S10	−0.31	0.24	0.08	0.42	−0.41	−0.30	−0.59	0.28	0.37	0.52	−0.45	0.42	−0.02	0.35
Method III	S1	0.68	−0.18	0.17	0.42	−0.60	0.04	−0.47	−0.20	−0.16	−0.20	0.83	−0.35	0.04	−0.83
	S2	−0.23	0.45	−0.33	−0.02	0.22	−0.31	−0.39	0.09	0.19	0.23	−0.28	0.58	−0.28	0.36
	S3	0.19	−0.02	0.52	0.17	0.04	0.18	0.21	−0.19	−0.16	−0.10	0.07	−0.18	0.75	−0.08
	S4	−0.57	0.06	−0.16	−0.39	0.55	−0.14	−0.33	0.35	−0.14	0.08	−0.76	0.18	−0.11	0.80
	S5	0.33	−0.09	−0.10	0.29	−0.25	−0.09	0.26	−0.13	−0.38	0.06	0.27	0.06	−0.14	−0.20
	S6	0.24	0.28	0.00	0.08	−0.18	−0.26	0.45	−0.51	−0.04	−0.01	0.30	−0.07	0.03	−0.38
	S7	0.57	−0.02	0.27	0.36	−0.46	0.23	0.64	−0.53	−0.27	−0.21	0.63	−0.42	0.48	−0.60
	S8	0.04	−0.25	−0.19	−0.10	0.22	−0.06	−0.31	0.15	0.28	0.57	−0.18	0.18	−0.28	−0.01
	S9	−0.31	0.18	−0.06	−0.38	0.34	−0.03	−0.52	0.40	0.56	0.36	−0.39	0.14	0.16	0.36
	S10	−0.29	0.21	−0.13	−0.54	0.29	−0.12	−0.51	0.53	0.35	0.29	−0.41	0.23	0.08	0.37

Correlations between traits
Method I = self report, Method II = interviewer report, Method III = observer report.

Table 1. *Continued*

S5	S6	S7	S8	S9	S10	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1.00															
0.56	1.00														
0.46	0.40	1.00													
−0.47	−0.61	−0.34	1.00												
−0.53	−0.55	−0.30	0.68	1.00											
−0.27	−0.39	−0.34	0.71	0.63	1.00										
0.45	0.10	0.47	−0.26	−0.31	−0.39	1.00									
0.07	0.10	−0.32	0.07	0.03	0.31	−0.40	1.00								
−0.12	−0.21	0.28	0.21	0.11	0.08	0.13	−0.35	1.00							
−0.34	−0.14	−0.56	0.18	0.08	0.24	−0.80	0.28	−0.18	1.00						
0.45	0.20	0.34	−0.41	−0.46	−0.30	0.36	−0.22	0.06	−0.21	1.00					
0.27	0.61	0.59	−0.32	−0.29	−0.25	0.29	−0.11	0.13	−0.38	0.31	1.00				
0.50	0.37	0.78	−0.26	−0.41	−0.32	0.56	−0.36	0.44	−0.57	0.36	0.42	1.00			
−0.06	−0.15	−0.13	0.34	0.47	0.37	−0.11	0.22	−0.24	0.06	0.02	−0.09	−0.27	1.00		
−0.42	−0.40	−0.53	0.55	0.67	0.64	−0.45	0.25	0.04	0.27	−0.64	−0.49	−0.45	0.27	1.00	
0.34	0.42	−0.51	0.37	0.56	0.58	−0.38	0.16	−0.17	0.35	−0.33	−0.48	−0.44	0.22	0.71	1.00

the exceptions. For items S5 and S6 the patient–interviewer correlations are poor, only 39% (S5) and 28% (S6) of the validity diagonal values are higher than the corresponding correlations in the heterotrait–heteromethod row and column. The patient–observer results are 61 and 78% respectively and insignificant.

Turning to the validity Criterion 3, Table 4 compares the validity diagonal values with the correlations in the heterotrait–heteromethod block. Generally one finds the same pattern as in Table 3, the performance of items S5, S6 and S8 is rather poor, whereas the rest of the items show evidence of a high degree of validity.

In Table 5 the pattern of trait intercorrelations for the mono- and heteromethod blocks is presented (Criterion 4). The majority of the items, seven out of nine, had an average number of correlations with the same sign above seven. Two of the items, S4 and S5 showed an average of 6.7 and 6.1 respectively.

For the two global psychosocial well-being questions a full multitrait–multimethod analysis was not undertaken. The validity diagonal values are presented (correlations between the same trait using different methods). The validity diagonal correlations are significant and reasonably high for both the two traits (Table 6). The happy/unhappy item

Table 2. Psychosocial well-being. Validity diagonal values—Criterion 1

Trait		Method I/II Patient/ interviewer	*P*	Method I/III Patient/ observer	*P*	Method II/III Interviewer/ observer	*P*
S1	Feeling of strength and energy	0.78	<0.0001	0.68	<0.0001	0.83	<0.0001
S2	Loneliness	0.43	<0.01	0.45	<0.01	0.58	<0.0001
S3	Relationship with other people	0.47	<0.01	0.52	<0.01	0.75	<0.0001
S4	Tiredness and exhaustion	0.70	<0.0001	0.55	<0.001	0.79	<0.0001
S5	Sense of satisfaction with oneself	0.23	0.104	0.29	<0.05	0.44	<0.05
S6	Feeling that life is worth living	−0.12	0.27	−0.26	0.08	0.60	<0.0001
S7	Feeling of happiness	0.73	<0.001	0.64	<0.0001	0.77	<0.0001
S8	Lack of confidence in oneself	0.38	<0.05	0.15	0.22	0.35	<0.05
S9	Feeling depressed	0.44	<0.01	0.56	<0.001	0.67	<0.0001
S10	Feeling of meaninglessness	0.52	<0.001	0.30	0.05	0.58	<0.0001

Table 3. Psychosocial well-being. Validity diagonal > heterotrait–heteromethod row and column—Criterion 2. Number of correlations meeting the criterion divided by number of correlations examined for the criterion presented as a percentage

Trait	% Patient/ interviewer	% Patient/ observer	% Interviewer/ observer
S1	100	100	100
S2	89	89	100
S3	100	100	100
S4	89	89	94
S5	39	61	89
S6	28	78	100
S7	100	100	100
S8	78	28	72
S9	89	100	100
S10	100	61	94

Table 4. Psychosocial well-being. Validity diagonal > heterotrait monomethod—Criterion 3. Number of correlations meeting the criterion divided by number of correlations examined for the criterion presented as a percentage

Trait	% Patient/ interviewer	% Patient/ observer	% Interviewer/ observer
S1	94	94	94
S2	100	100	100
S3	100	100	100
S4	94	89	94
S5	28	39	67
S6	17	50	100
S7	100	100	100
S8	61	33	72
S9	72	94	89
S10	94	50	89

shows a high degree of validity with correlations equal or above 0.53 for all three. For the satisfied/dissatisfied item the correlations are 0.61, 0.30 and 0.32, although still significant at the 5% level.

Medical symptoms

In Table 7 the validity diagonal values, three for each trait, are presented. Except for one of the correlations, patient–observer for trait M6 'sore and red skin' (0.17) all the correlations are statistically significant at the 5% level (Criterion 1). Table 8 shows the percentage of validity diagonal values which are higher than the values of the corresponding columns and rows in the heterotrait–heteromethod triangles (Criterion 2). Trait M6 again shows a low patient–observer value (55%). For the rest of the symptom measures, the results are satisfactory. Table 9 refers to the results of Criterion 3. This criterion is met to a very high extent for all symptom measures.

With the regard to Criterion 4 the questions were divided into three categories, i.e. chemotherapy-related (M1, M2 and M3), radiotherapy-related (M4, M5 and M6) and disease-related (M7, M8, M9 and M10) and analyzed separately. In Table 10 the trait intercorrelations for the mono- and heteromethod blocks are presented. A similar pattern of trait intercorrelations is shown for the chemotherapy-related items. For the radiotherapy-related items M4 and M5 a similar pattern is shown, with correlations varying from 0.47 to 0.71. With regard to item M6, the variation in correlations are higher, 0.09–0.83 (M4/M6) and 0.01–0.31 (M5/M6). For the majority of the disease-related items correlations were similar; however, item M9 did not show a similar pattern of trait intercorrelations. The correlations M8/M9 varies from 0.02 to 0.37 and M9/M10 from −0.25 to 0.18.

Table 5. Psychosocial well-being. Pattern of trait intercorrelations—Criterion 4. Percentage of correlations with the same sign

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1										
S2	9									
S3	9	9								
S4	7	7	7							
S5	7	5	5	7						
S6	9	5	7	7	5					
S7	8	9	9	7	7	9				
S8	7	8	6	5	7	9	9			
S9	9	9	5	6	7	9	9	8		
S10	9	9	6	7	5	7	9	9	9	
Average	8.2	7.8	7.0	6.7	6.1	7.4	8.4	7.6	7.8	

Table 6. Psychosocial well-being—global questions. Validity diagonal values. Correlation between the same traits using different methods

	Patient/ interviewer	*P*	Patient/ observer	*P*	Interviewer/ observer	*P*
Happy/unhappy	0.53	<0.001	0.60	<0.0001	0.63	<0.0001
Satisfied/dissatisfied	0.61	<0.0001	0.30	<0.05	0.32	<0.05

Table 7. Medical side-effects. Validity diagonal values—Criterion 1. Correlations between the same traits using different methods

Trait		Method I/II Patient/ interviewer	*P*	Method I/III Patient/ observer	*P*	Method II/III Interviewer/ observer	*P*
M1	Nausea	0.76	<0.0001	0.71	<0.0001	0.92	<0.0001
M2	Vomiting	0.75	<0.0001	0.72	<0.0001	0.79	<0.0001
M3	Hair loss	0.38	<0.05	0.43	<0.05	0.79	<0.0001
M4	Dysphagia	0.85	<0.0001	0.72	<0.0001	0.69	<0.0001
M5	Sore throat	0.83	<0.0001	0.76	<0.0001	0.83	<0.0001
M6	Sore and red skin	0.70	<0.0001	0.17	0.19	0.29	0.06
M7	Tired	0.74	<0.0001	0.68	<0.0001	0.70	<0.0001
M8	Appetite	0.84	<0.0001	0.80	<0.0001	0.90	<0.0001
M9	Sleep	0.62	<0.0001	0.47	<0.01	0.54	<0.001
M10	Pain	0.62	<0.0001	0.67	<0.0001	0.87	<0.0001

Activities of daily living

Table 11 displays the validity diagonal values for the items concerning activities of daily living scale. All correlations are above 0.53 and significant at the 5% level.

Physical performance

The physical performance items are analyzed by calculating the correlations between methods (Table 11). These items are supposed to make up a Guttman scale. For item F1, correlations can not be calculated due to the lack of variance in the responses. For item F2, the correlation between patient rating and interview rating is 0.47 and the correlation between interviewer and observer rating is 0.54. These correlations are significant at the 1% level. The correlations for Methods I and III are low (0.29) yet significant. Item F3 shows high correlations varying from 0.69 to 0.86, significant at the 5% level.

DISCUSSION

A person's subjective experience of his/her quality of life is not directly measurable. Indirect indicators are therefore necessary. In many cancer clinical trials the physical performance status and/or medical side-effects are the only indicators used. Such an approach may be too limited. A cancer patient often asks three questions: 1. What kind of treatments are available? 2. How long can I live? 3. What are the treatment-related side-effects and how do they influence my quality of life? Indicators of

Table 8. Medical side-effects. Validity diagonal > heterotrait–heteromethod row and column—Criterion 2. Number of correlations meeting the criterion divided by number of correlations examined for the criterion presented as a percentage

Trait	% Patient/ interviewer	% Patient/ observer	% Interviewer/ observer
M1	100	100	100
M2	100	100	100
M3	90	90	100
M4	100	100	100
M5	100	100	100
M6	100	55	95
M7	100	100	100
M8	100	100	100
M9	100	100	95
M10	100	100	100

Table 9. Medical side-effects. Validity diagonal > heterotrait monomethod row and column—Criterion 3. Number of correlations meeting the criterion divided by number of correlations examined for the criterion presented as a percentage

Trait	% Patient/ interviewer	% Patient/ observer	% Interviewer/ observer
M1	100	90	100
M2	100	90	95
M3	90	100	100
M4	100	100	100
M5	100	100	100
M6	100	80	90
M7	100	100	100
M8	100	100	100
M9	100	100	95
M10	100	100	100

psychosocial well-being seems therefore necessary in a more comprehensive assessment of quality of life. Several methods to obtain this kind of information have been described in the literature of psychosocial oncology. These include short questionnaires, long extensive questionnaires, diary cards, structured and open-ended interviews [22–24].

It seems highly desirable to use methods which can discriminate between the effects of various treatments on quality of life. These methods should also be applicable in a daily clinical setting and during follow-up. In addition, such methods should be easily understandable and not too time-consuming. The questionnaire presented in this study was based on these considerations.

Psychosocial well-being

The validity of the majority of the items in our questionnaire is high with exception for three items, S5 'have been pleased with yourself', S6 'have felt that life is worth living' and item S8 'have lacked confidence in yourself' (Tables 2–4). However, with regard to Criterion 4 both items S4 and S5 showed a low pattern of trait intercorrelations (Table 5).

The validity is doubtful because of a lack of agreement between the patient on the one hand, and the interviewer and observer on the other. The correlations between interview and observer are significant at the 5% level for all three items (0.44, 0.60 and 0.35) (Table 2).

This might be explained as a result of the considerable overlap between the two methods. Both interviewer and observer based their evaluations on information obtained in the same interview. The lack of agreement between patient and observer/ interviewer might have several explanations.

Correlations are higher when more general questions are asked. Low correlations appear when indicators deal with more personal matters such as items S5, S6 and S8.

The semi-structured interview may not have pro-

Table 10. Medical side-effects. Trait intercorrelations in mono- and heteromethod blocks—Criterion 4

Chemotherapy-related (M1, M2 and M3)			Radiotherapy-related (M4, M5 and M6)			Disease-related (M7, M8, M9 and M10)					
M1/M2	M1/M3	M2/M3	M4/M5	M4/M6	M5/M6	M7/M8	M7/M9	M7/M10	M8/M9	M8/M10	M9/M10
0.73	0.34	0.39	0.71	0.83	0.10	0.39	0.17	0.34	0.06	0.52	−0.08
0.55	0.54	0.53	0.65	0.29	0.17	0.37	0.28	0.39	0.11	0.37	−0.25
0.66	0.29	0.36	0.58	0.09	0.25	0.64	0.25	0.58	0.02	0.36	0.18
0.60	0.21	0.40	0.62	0.22	0.31	0.57	0.33	0.47	0.11	0.38	−0.18
0.63	0.56	0.17	0.71	0.39	0.10	0.36	0.35	0.45	0.32	0.31	−0.17
0.57	0.14	0.56	0.68	0.03	0.01	0.67	0.14	0.49	0.10	0.27	0.12
0.77	0.23	0.40	0.58	0.37	0.08	0.48	0.24	0.53	0.30	0.28	−0.20
0.65	0.23	0.30	0.47	0.16	0.25	0.62	0.25	0.39	0.14	0.30	−0.02
0.83	0.23	0.29	0.49	0.25	0.01	0.62	0.48	0.40	0.37	0.25	0.03

Table 11. *Activities of daily living (ADL) and physical performance. Validity diagonal values. Correlation between the same traits using different methods (validity diagonal values)*

Trait	Patient/ interviewer	P	Patient/ observer	P	Interviewer/ observer	P
Activities of daily living						
A1	1.0	<0.0001	0.53	<0.0001	0.53	<0.0001
A2	0.75	<0.0001	1.0	<0.0001	1.0	<0.0001
A3	1.0	<0.0001	0.60	<0.0001	0.80	<0.0001
A4	1.0	<0.0001	0.70	<0.0001	0.70	<0.0001
Physical performance						
F1	—		—		—	
F2	0.47	<0.01	0.29	<0.05	0.54	<0.001
F3	0.69	<0.001	0.75	<0.0001	0.86	<0.0001

vided flexibility for describing the patients' personal matters. It should be kept in mind that an interview lasting from 45 to 60 min is limited in its ability to penetrate deep personal problems. Furthermore, the wording of the questions might not be familar to the patients, perhaps being too academic and far from the daily way of thinking.

As a final point it may be important that both the interviewer and the observer are health professionals (i.e. nurses and physicians), and their knowledge of the disease is quite different from that of the patients. Knowledge of the seriousness of the disease may introduce some bias, making the interviewer and observer assessments more negative than the patients' own feelings. However, when data were analyzed by cross-tabulation, patterns supporting the above hypothesis were not found.

At present a definitive choice between these explanations is not possible. As far as the three items are concerned, it can be concluded that one or more of the methods compared has low validity. Whether this is due to lack of validity of the self-rating method, or the two sets of ratings based on semistructured interview, or possibly of all three types of measures, the present analysis is unable to tell.

The present method of validation is therefore somewhat unsatisfactory: there exists a validity problem, but one lacks more specific information regarding the precise location of the measurement problem. For the time being a reasonable strategy would be to concentrate on the remaining seven items when drawing substantive conclusions about the effects of various treatments on psychosocial well-being. In view of the preliminary character of the present results, including only 31 patients, the three problematic items with low validity, concerning personal matters, are included in future studies in order to get more information about the performance of these measures.

Medical side-effects

Selby [10] has validated several of the chemotherapy side-effects by comparing doctors' ratings and self-ratings. High correlations were found for two of the most common chemotherapy side-effects, nausea (0.86) and vomiting (0.85). However, correlations were rather low concerning hair loss (0.35) and lack of appetite (0.38). In our study, the correlation for hair loss, item M3, is below 0.50 for patient/interviewer and patient/observer comparisons. However, they are still significant at the 5% level. This pattern is thus similar to the results of Selby's study. The low correlation may be due to a misunderstanding of the questions by bald men, despite the fact that the questions ask specifically about the last 14 days. The interviewer/observer correlation is high (0.79), and supports this explanation. The validity of item (M6) 'sore and red skin', is low. The physician disagrees strongly with the patient responses, probably indicating that the wording of the item is poor. The item is intended to describe local acute radiotherapy side-effects of sore and red skin. Many patients may have taken the question to be more general, addressing all skin problems.

According to Criterion 4, item M9, 'sleeping problems', also seems to have a low validity; however, it displayed a high degree of validity for Criteria 1, 2 and 3.

Activities of daily living and physical performance

Both scales can be categorized as 'Guttman scales' [25], and should as such satisfy special scaleability criteria. A 'Guttman scale' is a scale constructed by a set of items with response choices representing a continuum of increasing levels of intensity, difficulty or severity. In this study, however, our aim was to compare different methods and use these results to

analyze validity. An analysis of the scaleability of the Guttman item is best suited in studying large numbers of patient populations. It will be done in a later study [26].

In assessing activities of daily living, the validity diagonal values were found significant. Concerning the physical performance scale, a correlational analysis was impossible due to the fact that response to one of the items showed no variation at all. The reason for this insufficient variance is due to the patient inclusion criteria which excluded patients below a specific status level. However, it is important to include the activity of daily living scale in studies covering a longer follow-up period, when variations in performance status are expected to be substantial.

Acknowledgements—We thank the physicians who listened to the taped interviews, Lars Borge, Stein Kvaløy, Terje Risberg, Einar Sandven, Stein Sundstrøm, Erik Thorud and Erik Wist, and to Guri Økseter Halvorsen for typing the manuscript.

REFERENCES

1. Schipper H. Measuring quality of life, risks and benefits. *Cancer Treat Rep* 1985, **69**, 1115–1125.
2. Geddes DM. Quality of life. *Clin in Oncol* 1985, **4**, 161–175.
3. Karnofsky D, Abelmann WH, Craver LF, Burchenal JH. The use of nitrogen mustards in the palliative treatment of carcinoma. *Cancer* 1948, 634–656.
4. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: *Evaluation of Chemotherapeutic Agents*. MacLeod, Columbia University Press, 1949, 191–205.
5. Hutchinson TA, Boyd NF, Feinstein AR. Scientific problems in clinical scales, as demonstrated in the Karnofsky index of performance status. *J Chron Dis* 1979, **32**, 661–666.
6. Milstein JM, Cohen ME, Sinks LF. The influence and reliability of neurologic assessment and Karnofsky performance score on prognosis. *Cancer* 1985, **56**, 1834–1836.
7. Selby PJ, Chapman J-AW, Etazadi-Amoli J, Dalley D, Boyd NF. The development of a method for assessing the quality of life of cancer patients. *Br J Cancer* 1984, **50**, 13–22.
8. *Incidence of Cancer in Norway* 1984, The Cancer Registry of Norway.
9. Aisner J, Hansen HH. Commentary: current status of chemotherapy for non-small cell lung cancer. *Cancer Treat Rep* 1981, **65**, 979–986.
10. Høst H, Thorud E, Hatlevoll R *et al.* A randomized study of radiotherapy versus chemotherapy with cisplatin and etoposide in non-small cell carcinoma of the lung. *Cancer Treat Symp* 1985, **2**, 101–103.
11. Kamman R, Christie D, Irwin R, Dixon G. Properties of an inventory to measure happiness (and psychological health). *New Zealand Psychol* 1979, **8**, 1–12.
12. Naess S. *Concepts and Methods in a Norwegian Study of Quality of Life*. Institute of Applied Social Research, 1979.
13. Fox RM, Tattersall MHN. On the receiving end—patient perception of the side effects of cancer chemotherapy. *Eur J Cancer Clin Oncol* 1983, **19**, 202–208.
14. Klastersky J. Combination chemotherapy with cisplatin, etoposide and vincristine in non-small cell lung carcinoma: a clinical trial of the EORTC Lung Cancer Working Party. *Cancer Treat Rep* 1983, **67**, 727–730.
15. Einhorn L. Random prospective study of cyclophosphamide, doxorubicin, and methotrexate (CAM) combination chemotherapy versus single-agent sequential chemotherapy in non-small cell lung cancer. *Cancer Treat Rep* 1982, **66**, 2005–2011.
16. Aaronson NK, Bakker W, Stewart AL *et al.* A multidimensional approach to the measurement of quality of life in lung cancer clinical trials. IV World Conference on Lung Cancer. Abstract, 1985.
17. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychol Bull* 1959, **56**, 81–105.
18. Kim JO, Muller CW. *Factor Analysis*. Beverley Hills, Sage Publications, 1978.
19. Levine MS. *Canonical Analysis and Factor Comparison*. Beverley Hills, Sage Publications, 1977.
20. Rummel RJ. *Applied Factor Analysis*. Evanston, Northwestern University Press, 1970.
21. Zeller R, Carnines EG. *Measurement in Social Sciences*. Cambridge, Cambridge University Press, 1980, 56–69.
22. Van Damm FSA, Linssen CAG, Couzijn AL. Evaluating 'quality of life' in cancer clinical trials. In: Buyse ME, Staquet M, Sylvester RF, *Cancer Clinical Trials Methdos and Practice*. Oxford Medical Publication, 1984, 26–43.
23. Fayers PM, Jones DR. Measuring and analysing quality of life in cancer clinical trials: a review. *Stat Med* 1983, **2**, 429–446.
24. Spitzer WO, Dobson AJ, Hall J *et al.* Measuring the quality of life of cancer patients. *J Chron Dis* 1981, **34**, 585–597.
25. Guttman L. A basis for scaling quantitative data. *Am Sociol Rev* 1944, **9**, 139–150.
26. Kaasa S, Mastekaasa A, Thorud E. Medical side-effects, physical performance and activity of daily living evaluated in a randomized controlled trial for patients with inoperable non-small cell lung cancer, limited disease (combination chemotherapy versus radiotherapy). In preparation.