Eur J Cancer, Vol. 29A, Suppl. 1, pp. S63–S69, 1993
Printed in Great Britain

0964–1947/93 $6.00+0.00

Reliability and Validity of a Quality of Life Questionnaire in Cancer Patients

Enzo Ballatori, Fausto Roila, Carlo Basurto, Sergio Bracarda, Mara Picciafuoco, Marcello Soldani, Lucio Crinò, Maurizio Tonato, Gabriella Milella and Albano Del Favero

Two studies were sequentially conducted to validate a new questionnaire which takes into consideration the most important variables which could influence quality of life evaluation. Particular attention was given to the methodology employed to collect data and to the patients' characteristics. In the first study 80 consecutive cancer patients were randomised to twice fill in one of four different types of questionnaire, each one characterised by a different polarisation of semantic and syntactic extreme values of the visual linear analogue (for instance, "very much" always on the right, regardless of the semantic value of the answer; positive semantic value always on the right, regardless of whether it was "very much" or "not at all"; and so on). The second study, conducted on 60 lung cancer patients, consisted in testing the reliability (by measuring the reproducibility in different ways) and the validity (by performing a factor analysis) of the type of questionnaire indicated by the first study as the most reliable. The internal coherence was also evaluated by measuring the effects of physical and psychological conditions on the responses.

Eur J Cancer, Vol. 29A, Suppl. 1, pp. S63–S69, 1993.

INTRODUCTION

THE INTEREST of the scientific community in the supportive care of cancer patients has constantly increased during the last few years. In fact, adequate supportive care has been demonstrated to improve the survival of cancer patients as well as their quality of life by preventing or reducing the severity of side-effects induced by cancer chemotherapy and relieving the symptoms due to the neoplastic disease itself. The evaluation of quality of life is therefore an important issue and deserves attention in a symposium dedicated to the supportive care of cancer patients. Despite its importance, there are still many uncertainties regarding the best way to evaluate quality of life [1–3]. Today, there is general agreement on the following points:

(1) We need to measure quality of life of cancer patients and its variations, possibly in relation to the clinical evolution of the disease and to the treatments administered to the patients.
(2) We must use measurement scales, such as visual linear analogue scales (VAS) or categorical scales (scores). The former seem preferable to categorical scales from the point of view of analysis of data (they have associated continuous variables instead of the discreteness of categorical scales), although more difficult to fill in.
(3) We need multidimensional measurements, as quality of life involves different aspects of the cancer patient's life, such as feeling, psychological condition, social relationships, patient's relationship with his or her disease and treatment toxicity.
(4) The evaluation must be performed by the patient; no other person can have an exact perception of the patient's quality of life as it depends on his or her feelings, intelligence, philosophy of life, perceptions, and so on.
(5) Results of the evaluation must be expressed in a very simple way, so as to use them easily in statistical analysis in combination with survival data, treatment response and therapy toxicity.
(6) Any new questionnaire should be tested for reproducibility and validity. Many of the available questionnaires do not meet the above-mentioned criteria and therefore the results of studies on quality of life which employ them are open to criticism.

Because of the lack of a general theory about validity and reproducibility of questionnaires, many empirical studies involving methodological aspects of the tools in use must be performed. Are we really sure that a quality of life questionnaire is measuring what we wish to measure? Do the results depend on the patient's feelings or on his or her real condition or perhaps on a systematic error which could be made in answering? Only empirical studies can give answers to these questions. Their results could be perceived as technical exercises, but we are convinced of their importance: too many quality of life studies have failed because such an assessment of validity and reproducibility has not been adequately carried out.

These considerations prompted us to plan a study with the aim of evaluating the reliability and validity of a new

---

Correspondence to E. Ballatori.
E. Ballatori and G. Milella are at the Medical Statistics Unit (Institute of Internal and Vascular Medicine), University of Perugia; F. Roila, C. Basurto, S. Bracarda, M. Picciafuoco, M. Soldani, L. Crinò and M. Tonato are at the Medical Oncology Division, Policlinico, Perugia; and A. Del Favero is at the Institute of Internal Medicine I, University of Perugia, 06100 Perugia, Italy.
Received 30 Jan. 1992; accepted 8 May 1992.

questionnaire on quality of life developed according to the above-mentioned guidelines.

The evaluation/validation process was carried out by conducting two consecutive studies. The aim of the first study was to identify the best way to arrange the various items included in the questionnaire. The aim of the second study was to evaluate the reliability and validity of the questionnaire in a group of lung cancer patients.

MATERIALS AND METHODS

In the first study 80 consecutive cancer patients were randomised to fill in one of four different types of questionnaires, A, B, C and D, each including 49 items covering the most important issues which could affect their quality of life: fatigue and malaise, psychological distress, sense of well-being, relationship with the hospital, change in abilities to earn, symptoms of lung cancer, side-effects of treatment, functional status, social interaction, opinion about illness, unidimensional measurement of quality of life.

After oral consent and following a brief explanation of the questionnaire and how it should be filled in, it was given to the patients twice (with a 24-h interval). The patient had to answer each question by marking a VAS 100 mm long with the extremes labelled with the words defining the range of answer. To avoid a carry-over effect, the questions in the patient's two questionnaires were distributed at random. In the 24-h interval between test and retest, a physician filled in his or her own questionnaire based on direct interview with the patient. As the main aim of this study consisted in evaluating if a particular polarisation of the VAS could improve the quality of the responses, the four types of questionnaires were defined by a different polarisation of semantic and syntactic extreme values of the VAS. More precisely:

Questionnaire A (A-q): "very much" was always on the right and "not at all" always on the left of the VAS, regardless of the semantic value of the response;

Questionnaire B (B-q): "not at all" and "very much" were assigned in random order to the left or to the right of the VAS, regardless of the semantic value of the response;

Questionnaire C (C-q): positive semantic value was always on the right of the VAS, regardless of whether it was "very much" or "not at all"; and

Questionnaire D (D-q): positive semantic value was always on the left of the VAS.

All measurements were expressed in terms of distance of the point marked by the patient from the extreme expressing the negative semantic value.

The second study consisted of testing the reliability and the validity of C-q enriched by four specific cancer chemotherapy side-effect items in 60 consecutive lung cancer patients.

In both studies the modalities of filling in the questionnaire were kept the same. The characteristics of the patients (sex, age, level of schooling, type of neoplasia, and so on) as well as an evaluation of ECOG performance status (PS) were recorded. Also in both studies the systematic errors were analysed. We define systematic error (briefly "error") as any response, given by the same patient, having a relevant discordance from test to retest within 24-h. More precisely, there was an error when in the test a score value was not lower than 70 or not greater than 30 and, for the same item, in the retest the score value was found, respectively, not to be greater than 30 or lower than 70.

A second type of mistake, less important than an error, was also considered: a missing value was defined as the lack of response to an item in at least one of the two questionnaires.

STATISTICAL METHODS

A two-tailed *t*-test for paired data was used, in the first study, to compare the mean values of the responses given by the patients to each item from test to retest 24 h later for each type of questionnaire. In the second study it was employed to detect the significant differences between the mean values of the responses to the same items given by the physician and by the patient. To complete the reliability assessment on the same data a Bravais–Pearson correlation coefficient was also calculated.

In both studies errors and missing values were analysed with respect to the patients and to their characteristics, by determining the percentages of patients with at least one error (and one missing value) as well as the mean number of errors (and of missing values) for patient and the corresponding concentration ratio, *R*.

The concentration ratio, *R*, is a normalised statistical index, therefore varying between 0 and 1. Its highest values indicate that the errors (or the missing values) are concentrated in a few people; instead, *R* values near 0 indicate the errors (or missing values) are rather uniformly distributed among all people.

This analysis was performed in the first study not only to show which types of questionnaires could produce a smaller number of errors and of missing values, but also to detect which characteristics of patients could increase the mean number of errors and missing values. Because of the imbalance of these characteristics among the groups defined by type of questionnaire, a multifactorial analysis based on log–linear models was needed to adjust the effect of each factor on the other ones.

Moreover, in the first study the frequency distributions of errors (and of missing values) were also investigated with respect to the items where they were found. After a descriptive analysis mainly based on the mean values and on concentration ratio, a Poisson distribution was used to investigate the variability of errors (and of missing values) among the items.

To validate the questionnaire, in the second study, the construct validity was analysed and a factor analysis was performed.

The construct validity [3] of the questionnaire can be analysed by determining if the relation between the major variables is consistent with our prior expectation and which item synthetised in the best way the underlying concept of which levels we need to measure in the patients.

On these bases, and after classifying the 53 items into 11 groups defined by quality of life dimensions, as previously indicated, we evaluated the dependence of all the responses on the physical conditions and on two items whose response expressed psychological distress of the patient with respect to his or her own disease.

As a working hypothesis we assumed that:

(i) the patient physical condition, evaluated by the ECOG PS, could especially modify his or her responses to items concerned with his or her functional status;

(ii) the patient psychological condition, indicated by the patient opinion about his or her illness, could especially influence his or her responses to items concerned with his or her psychological sphere (psychological distress, social interaction, sense of well-being).

As the extreme scores of any item are the most informative, we indicated by L a very low score (lower than 30), i.e. a particular distress with respect to the issue requested by the item, and by H a very high score (greater than 70), i.e. a very good fit of the patient with respect to question content. To test the first working hypothesis, for each item, the percentages of patients who chose L (and, separately, H) were compared between two classes of ECOG PS ($E_0=0$ and $E_1=1, 2, 3, 4$) and a χ^2-test for independence was performed. To test the second working hypothesis we relied on the patient's opinion about his or her illness (questions 35 and 36) and therefore we classified the patients in S_1, S_0 if, respectively, they believed their disease was very serious (response to item 36 lower than 30) or not so very serious (score not lower than 30). Then the percentages of patients who chose an extreme score (L and H separately) of any item were compared between S_0, S_1 groups and a χ^2-test for independence was performed. Obviously the prior expectations were that the percentages of patients choosing L would be greater in the S_1 than in the S_0 group, as to enhance the pessimistic point of view of the distressed subjects.

In the same way, with respect to item 35, the patients were classified into C_1, C_0 groups if, respectively, they believed their disease was very difficult to cure (score lower than 30) or not so difficult to cure (score greater than 70).

Then the percentages of patients who chose an extreme score (L and H, separately) of any item were compared between C_1, C_0 groups and a χ^2-test was performed.

Finally, a factor analysis, conducted on the results obtained in the first questionnaire filled out by the patients, was also performed to detect the main factors which explained the greater part of variability of responses.

RESULTS OF THE FIRST STUDY

Comparing the mean values of the responses to each item between two questionnaires

To detect, for each type of questionnaire, if the mean value of the responses to each item had changed between the test and the retest within 24 h, a two-tailed t-test for paired data was used with a significance level ($P \leqslant 0.1$), chosen so high because of the low number of patients in each group.

49 items were analysed and the number of them with a significant difference between the mean values for type of questionnaire is nine for A-q, eight for B-q, six for both C-q and D-q. On the bases of the chosen significance level, under the null hypothesis of equality of the mean values, the expected number of significant differences for groups, i.e. those uniquely due to chance, is about five. Therefore it seems that, with respect to this criterion, all four types of questionnaires were satisfactory, especially D-q and C-q.

Systematic measurement errors and missing values: analysis of patients

4 patients were excluded from this analysis because they did not fill in one or more pages of at least one of the two questionnaires.

The percentages of patients with at least one error (at least one missing value) are very high for all types of questionnaires: 72.2% (50.0%) for A-q, 76.2% (50.0%) for B-q, 84.2% (52.6%) for C-q, 83.2% (22.2%) for D-q.

Table 1. Mean number of errors and missing values for subject and corresponding concentration ratio (R) for type of questionnaire

	A-q	B-q	C-q	D-q
Errors				
Mean number	2.72	2.90	2.79	2.11
R	0.574	0.537	0.475	0.396
Missing value				
Mean number	1.06	1.10	3.00	0.50
R	0.693	0.632	0.754	0.791

Instead, considering the mean number of errors and of missing values (Table 1), D-q seems to have the best results. But these different performances of the four types of questionnaires could be due to their intrinsic structure or to differences in patient characteristics which could influence the results.

Therefore, an analysis in this direction has been performed. The results identify the most important patient characteristics able to influence the questionnaire performance, such as level of instruction, age and ECOG PS. In fact, the mean number of errors (of missing values), for subject, is lower for the younger people (<55 years old) than for the older people ($\geqslant$55 years old): 1.72 vs. 3.12 (0.59 vs. 1.95); for the people with a level of instruction higher than elementary school: 1.73 vs. 2.98 (0.73 vs. 1.74); for the patients in better (ECOG = 0) rather than in worse condition (ECOG $\geqslant$ 1): 1.39 vs. 2.32 (0.81 vs. 1.63).

These characteristics do affect the responses and they are not well balanced among the four groups of patients, making C-q the most unfavourable and D-q the most favourable: 14/19 are more than 54 years old (9/18 for A-q, 12/21 for B-q, 7/18 for D-q); 12/19 have the lowest level of instruction (14/18 for A-q, 10/21 for B-q, 10/18 for D-q); 14/19 are in the worst (ECOG $\geqslant$ 1) physical condition (12/18 for A-q, 11/21 for B-q, 12/18 for D-q).

At a multifactorial analysis conducted by log-linear models (Table 2) the mean number of errors resulted significantly superior for the patients with the lowest level of instruction

Table 2. Log-linear models for the number of errors and missing values for patient. Two-tailed z-test for significance of linearly independent contrasts between levels of the factors

	Errors		Missing values	
Factors and contrasts	z-test	Significance level	z-test	Significance level
Level of inst. *				
L–H	2.09	0.037	2.44	0.015
Age†				
O–Y	2.81	0.005	2.44	0.015
ECOG PS‡				
E_0–E_1	−1.24	n.s.	−0.98	n.s.
Type of questionnaire				
(D-q)–(C-q)	−0.34	n.s.	−4.44	0.001
(A-q)–(C-q)	0.35	n.s.	−3.65	0.001
(B-q)–(C-q)	−0.14	n.s.	−3.19	0.002

*L = at most, elementary school; H = superior.
†Y = <55 years old; O = $\geqslant$55 years old.
‡$E_0=0$; $E_1=1, 2, 3, 4$.
Remark: G-test (equal to minus two times maximum likelihood ratio) is significant if and only if the correspondent z-test has reached a significant value; for this reason G-tests are not shown.

and for the oldest patients: neither the ECOG PS nor the type of questionnaire adjusted for the above-mentioned factors showed a significant effect. Instead, the mean number of missing values resulted significantly superior for C-q.

As concerns the concentration ratio (R) values, they are higher for the missing values than for the errors. By analysing the parameter values needed to calculate R, we can outline for the errors:

A-q: 5.6% of patients are responsible for 30.6% of the errors;

B-q: 4.8% of patients are responsible for 21.3% of the errors;

C-q: 5.3% of patients are responsible for 17.0% of the errors;

D-q: 5.6% of patients are responsible for 18.4% of the errors.

Instead, for the missing values we can see that:

A-q: 22.2% of patients are responsible for 73.7% of missing values;

B-q: 35.0% of patients are responsible for 86.4% of missing values;

C-q: 21.1% of patients are responsible for 82.5% of missing values;

D-q: 22.2% of patients are responsible for 100% of missing values.

Systematic measurement errors and missing values: analysis of items.

The mean number of errors per item shows that the errors are much more frequent in B-q (1.24) and in A-q (1.08) than in C-q (0.69) and in D-q (0.37). Instead, the mean number of missing values per item is greater in C-q (1.22) than in A-q, B-q (both 0.41) and in D-q (0.12). Moreover, both the missing values and the errors are concentrated in a few items, particularly the former ones. Therefore, it would be sufficient to remove one or a few items to obtain a substantial decrease in the mean number of both errors and missing values.

Finally, for all four types of questionnaire, the number of errors as well as the number of missing values for each item are very well approximated by a Poisson process. In fact, the χ^2-test for homogeneity between empirical and theoretical distribution never reaches a significant value. Therefore, it seems that the errors and the missing values are independent of the type of items: their presence seems to be due purely to a random effect.

On the contrary, the number of errors and missing values for each subject do not follow a Poisson process: there is a tendency to make errors more than once on some subjects; errors and missing values are not independent of patients.

Comparing the reliability

For each item and for each type of questionnaire, a Bravais–Pearson correlation coefficient was calculated between the response given in test and retest within 24 h and the values obtained were aggregated in three classes:

Very good agreement: $r \geqslant 0.8$.

Good agreement: $0.6 \leqslant r < 0.8$.

Poor agreement: $r < 0.6$.

D-q seems to have the best performance: for 27 items there is very good agreement (VG), for 17 items good agreement (G) and only for 5 items poor agreement (P). C-q follows: 19 VG, 20 G, 10 P. The worst results were achieved by B-q (17 VG, 16 G, 16 P) and A-q (14 VG, 12 G, 23 P). Moreover, when considering all responses with at least good agreement ($r \geqslant 0.6$), 44 items had good reliability in D-q, 39 in C-q, 33 in B-q and 26 in A-q.

Discussion

There are very few empirical studies about the methodological aspects of measurement scales used in social research; our study should be a contribution in this field. Unfortunately, the randomisation of patients to the four types of questionnaires did not yield a good balance of the characteristics of subjects, and seriously affected the comparison between types of questionnaires. C-q was found particularly unfavourable for the relatively more frequent presence of old patients, with a low level of instruction and with a high ECOG score; instead, D-q resulted particularly favourable for the opposite composition of patient characteristics.

In conclusion, while it is reasonable not to recommend the use of A-q and B-q, the choice between C-q and D-q is uncertain: it seems that a uniformly semantic polarisation of the responses is the best way to use a linear VAS to measure quality of life.

RESULTS OF THE SECOND STUDY

60 consecutive lung cancer patients (54 males, 6 females) filled out C-q (semantic positive polarisation on the right). 3 males left blank one or more pages of at least one questionnaire and therefore were excluded from the analyses. The patient group was more homogeneous with respect to relevant patient characteristics than that considered in the previous study: 33/57 patients were 50–64 years old, 44/57 had the lowest level of instruction (at most elementary school), 41/57 had an ECOG PS not lower than 1. Perhaps better information than that furnished in the first study led to an inferior mean number of errors per subject (1.44 vs. 2.64) and of missing values (0.93 vs. 1.42). Other important characteristics are: 35/57 had non-small cell cancer and only 9/57 were in disease progression.

Self-evaluation and physician's recording comparison

For each item, the mean values of the response given by the patient and the physician were compared by t-test for paired data. A significant difference was found in 46/53 items for the test and in 45/53 for the retest within 24 h. On the contrary, there were significant differences between patients' answers at test and retest in only 8 items.

The poor performance of the physician in evaluating quality of life is also shown by measuring the correlation between physician's answers to those of the patients (Table 3). As can easily be seen, physicians are not able to reliably judge a patient's psychological condition. Therefore, the physician's

Table 3. Number of responses with respect to the agreement measured by Brevais–Pearson's correlation coefficient (lung cancer patients)

Degree of agreement	Physician /test	Physician/retest within 24 h	Patient test/retest
Very good ($r \geqslant 0.75$)	3	5	17
Good ($0.5 \leqslant r < 0.75$)	29	34	32
Poor ($r < 0.5$)	21	14	4

Table 4. Percentages of patients choosing L score (0–30) and H score (70–100) for the functional status items for classes of ECOG PS ($E_0 = 0$ *and* $E_1 = 1, 2, 3, 4$)

	L score		H score	
Functional status items	E_0	E_1	E_0	E_1
13—How much does pain limit you in your daily activity?	12.5	31.7	56.3	34.2
24—Are you able to take care of yourself?	0.0	14.6	81.3	61.0
25—Are you able to do light work around the house?	18.8	43.9	75.0	14.6*
27—Are you able to perform your job?	25.0	63.4*	50.0	9.8*
28—Do you exercise regularly?	31.3	24.4	37.5	12.2*

*$P \leqslant 0.05$, with respect to χ^2-test.

evaluation was no longer considered useful in validating the questionnaire.

Errors and missing values

Although not significant, all the differences in the mean number of errors and in missing values between the factors levels are ordered in the same way as in the previous study: all mistake indicators are greater in older patients, in patients with a low level of instruction and in those with progressive disease. Moreover, the mean number of errors and of missing values are greater among the patients with the worst performance status.

Finally, it can be shown that the presence or not of errors is independent of missing values. In fact, among the 40 patients who made no errors, 14 also had no missing values (35%) and among the 17 patients with at least one error, 6 (35.3%) had no missing value.

Validation of the questionnaire

The construct validity of the questionnaire was examined both to provide tests for the internal coherence and to perform a factor analysis to synthetise the underlying concepts whose levels need to be measured in the patients.

Table 5. Percentages of patients choosing L score (0–30) and H score (70–100) for some items belonging to the psychological sphere for classes of ECOG PS ($E_0 = 0$ *and* $E_1 = 1, 2, 3, 4$)

	L score		H score	
Functional status items	E_0	E_1	E_0	E_1
03—Do you feel discouraged?	6.3	34.2*	56.3	24.4*
04—Do you feel sad?	18.8	26.8	43.8	17.1*
05—Do you want to cry?	6.3	22.0*	87.5	53.7*
07—Do you feel uncertain and undecided even about unimportant things?	0.0	22.0*	62.5	39.0
17—How much has your illness changed your relationship with your friends?	6.3	31.7*	75.0	39.0*
38—How much time do you spend thinking about your illness?	18.8	36.3*	43.8	14.6*

* $P \leqslant 0.05$, with respect to χ^2-test.

As tests of internal coherence, the effects of physical and psychological condition on the responses are outlined, based on the classification of 53 items divided into 11 quality of life dimension groups, as previously defined.

Effect of physical condition on the responses. L indicated a very low score (lower than 30), i.e. a particular distress with respect to the issue requested by the item, and H a very high score (greater than 70), i.e. a very good fit of the patient with respect to the content of the question.

If we compare, for each item, the percentage of patients who chose L (and, separately, H) between the two classes of ECOG PS ($E_0 = O$ and $E_1 = 1, 2, 3, 4$) we can show the evident relationship of the responses to items concerning the functional status on ECOG PS (Table 4).

The different ECOG PS significantly affects not only the answers to items belonging to the other dimensions related to physical condition (symptoms of lung cancer, side-effect of treatment, fatigue and malaise, relationship with the hospital, sense of well-being), but also the answers to many items concerned more specifically with the psychological sphere of the patient (mainly, social interaction and psychological distress) (Table 5).

It is therefore evident that physical impairment has strong repercussions in the patients psychological sphere, mainly when there is insufficient time to adapt to the altered physical condition.

Effect of opinion about the illness on the responses. Only 12/57 patients believed that their disease "can be cured with great difficulty" (C_1) and 20/57 patients that it "is very serious" (S_1).

The percentages of patients with these opinions decrease, although not significantly, with age and with level of instruction, while they increase with the ECOG PS. No patient with

Table 6. Percentages of patients choosing L score (0–30) and H score (70–100) for the psychological distress items, for classes defined by the opinion about illness (S_1 = *the disease is very serious;* S_0 = *the disease is not so very serious*)

Items belonging to "Psychological distress" area	L score		H score	
	S_1	S_0	S_1	S_0
03—Do you feel discouraged?	50.0	13.9*	15.0	41.7*
04—Do you feel sad?	45.0	16.7*	5.0	36.1*
05—Do you want to cry?	30.0	11.1	40.0	72.2*
06—Do you feel lonely?	20.0	8.3*	35.0	63.9*
07—Do you feel uncertain and undecided even about unimportant things?	35.0	5.6*	30.0	52.8
09—How irritable are you?	36.8	33.3	21.0	31.6
11—Do you forget things?	40.0	11.1*	40.0	50.0
31—Does illness influence your physical appearance?	63.2	33.3*	15.8	22.2
37—Are you worried about your illness?	80.0	44.4*	10.0	8.3
38—How much time do you spend thinking about your illness?	65.0	25.0*	5.0	33.3*
39—Do you believe in your treatment?	5.0	0.0	75.0	86.1

* $P \leqslant 0.05$, with respect to χ^2-test.

Table 7. *Percentages of patients choosing L score (0–30) and H score (70–100) for each item with a significant difference between proportions for classes defined by the opinion about illness (S_1 = the disease is very serious; S_0 = the disease is not so very serious)*

	L score		H score	
Areas and items	S_1	S_0	S_1	S_0
Fatigue and malaise				
01—How do you feel?	45.0	13.9*	0.0	30.6*
02—Do you feel tired?	70.0	19.4*	0.0	30.6*
Sense of well-being/satisfaction				
08—Do you have self-confidence?	25.0	2.8*	45.0	83.0*
14—How much are you able to have fun?	65.0	25.0*	0.0	8.3
21—Are you satisfied with the life you lead?	65.0	11.1*	5.0	50.0*
22—Do you feel useful?	55.0	8.3*	15.0	55.6*
Functional status				
13—How much does pain limit you in your daily activities?	42.1	13.9*	15.8	55.6*
24—Are you able to take care of yourself?	35.0	0.0*	50.0	75.0
25—Are you able to do light work around the house (repairs, gardening, etc)?	60.0	30.6*	10.0	41.7*
27—Are you able to perform your job?	84.2	29.4*	0.0	35.3*
28—Do you exercise regularly?	45.0	16.7*	5.0	27.8*
Social interaction				
17—How much has your illness changed your relationship with your friends?	47.4	11.1*	31.6	58.3
18—How much do you want to see people?	40.0	19.4	20.0	47.2*
20—How do you feel when you are with people?	20.0	5.6	20.0	61.1*
Uniscale				
23—How is the quality of your life?	35.0	2.8*	15.0	33.3
Symptoms of lung cancer				
12—Do you have pain?	40.0	11.1*	30.0	69.4*
18—How much do you eat?	35.0	8.3*	10.0	22.2
49—Has your appetite increased?	65.0	27.8*	5.0	19.4
50—Has your appetite decreased?	50.0	13.9*	25.0	52.8*
52—Are you constipated?	40.0	11.1*	20.0	50.0*
Side-effects of treatment:				
51—How much do you vomit?	21.1	2.8*	63.2	91.7*

* $P \leqslant 0.05$, with respect to χ^2-test.

the highest level of instruction believes his or her disease can be cured with great difficulty (0/7).

The dependence of the responses to questions about psychological distress on S_0, S_1 groups is evident from Table 6. For instance, the 50% of patients who believed their disease is very serious (S_1 group) feel very discouraged (lower score—lesser than 30—chosen for 03 item), while the same proportion is 13.9% among the patients who believe their disease is not so very serious (S_0 group).

Table 8. *Pearson's correlation coefficients between the responses and the extracted factors*

Questions	Factor 1	Factor 2	Factor 3
Do you feel discouraged?	0.843*	0.215	0.130
Do you feel sad?	0.831*	0.294	0.094
Do you feel tired?	0.323*	0.106	0.061
How do you feel?	0.819*	0.130	0.120
Pain limitation of daily activity	0.768*	0.120	0.256
Do you want to cry?	0.727*	0.051	0.337
Are you able to take care of yourself?	0.717*	0.089	0.085
Do you feel useful?	0.693*	0.254	0.140
Opinion about seriousness of your disease	0.638*	0.180	0.153
Do you have self-confidence?	0.609*	0.142	0.056
Do you have pain?	0.609*	0.106	0.335
Opinion about curability	0.584*	0.113	0.083
How much do you want to see people?	0.581*	0.295	0.098
Time spent thinking about illness	0.568*	0.379	0.321
Are you worried about illness?	0.547*	0.450	0.115
Satisfaction of sexual desires	0.538*	0.300	0.240
Change in friendship because of illness	0.534*	0.290	0.025
Time spent with friends	0.514*	0.348	0.034
How much do you want to make love?	0.437*	0.238	0.212
Problems for going to hospital	0.175	0.681*	0.266
Problems for hospital treatment	0.109	−0.648*	0.231
Problems for time spent in hospital	0.125	0.589*	0.253
Has your appetite increased?	0.424	0.158	−0.707*
How much do you eat?	0.528	0.073	0.625*
Has your appetite decreased?	0.556	0.195	0.548*

Factor 1: Psychological dimension of quality of life.
Factor 2: Relationship with the hospital.
Factor 3: Appetite modification.

Other items are also in significant relationship to the discouragement produced by pessimistic opinions about the illness, such as those items that referred to fatigue and malaise, functional status, symptoms, side-effects of treatment, social interaction and sense of well-being (Table 7).

By executing the same analysis for item 35, and classifying the patients in C_1, C_0 groups if, respectively, they believed their disease could be cured only with great difficulty, or was not so difficult to cure, analogous results are obtained. In conclusion, internal coherence of the questionnaire seems to be sufficiently proven. The chosen indicators show, however, a low specificity due to an interaction between the psychological and physical sphere, much more profound than our prior expectations.

Factor analysis. To validate the questionnaire, a factor analysis was also performed to detect the main factors which explained the greatest variability of responses.

The symptoms of lung cancer and the side-effects of treatment were no longer considered for their low correlation among themselves and with main factors mentioned above: one factor, and only one, was needed to explain the variability of each one of these responses. After many steps, using the Varimax rotation, a model with 25 items was found (Table 8).

The first factor explained about 50% of the variability of all the responses and it mainly pertained to those questions regarding the psychological sphere. The other two relevant factors were related to the relationship of the patient with the hospital and to appetite modification. These three factors globally explained about 75% of the variability of all 25 responses. In conclusion, the prior expectation to extract one factor for each area of the questionnaire vanished, probably due to the strict relationship between physical and psychological condition.

CONCLUSIONS

The most relevant conclusions which can be derived from both studies seem to be the following:

(1) we need a self-evaluated questionnaire;
(2) to avoid errors and missing values, we must be very clear in explaining to the patients the aims of the questionnaire and those technical aspects needed to fill it in correctly, especially to the oldest patients and to those with a low level of instruction and/or with a low performance status;
(3) using VAS, it is preferable that the semantic polarisation be the same for all the items;
(4) there is a strong interaction between physical condition and the psychological sphere of the patients: functional status is closely related to all social and psychological dimensions of life. Therefore, the questionnaire, though divided in sections, must regard the patient in his or her entirety;
(5) the patient opinion about his or her illness has relevant effects not only on psychological distress, but also on social interaction, functional status, sense of well-being; we think they must be carefully evaluated in every quality of life questionnaire;
(6) factor analysis could be used not only to validate new questionnaires, but also to synthetise the responses for each patient in a score so as to rank them in terms of standardised factor levels.

We are far from a satisfactory definition of a questionnaire able to measure quality of life of cancer patients, because in both the social and psychological sciences there are not yet universally accepted definitions of the areas of our interest. Empirical studies do not exist about reliability and validity of VAS and of categorical scales, synthetised in a general theory, mainly as concerns the errors which can be committed in using them: much space is available for social and psychological research in clinical oncology.

---

1. McMillen Moinpour C, Feigl P, Metch B, Hayden KA, Meyskens FL, Crowley J. Quality of life end points in cancer clinical trials: review and recommendations. *J Natl Cancer Inst* 1989, **81**, 485–495.
2. Donovan K, Sanson-Fisher RW, Redman S. Measuring quality of life in cancer patients. *J Clin Oncol* 1989, 7, 959–968.
3. Jones DR, Fayers PM, Simons J. Measuring and analyzing quality of life in cancer clinical trials: a review. In: Aaronson NK, Beckmann J, eds. *The Quality of Life of Cancer Patients*. New York, Raven Press, 1987, 41–51.