Pergamon

European Journal of Cancer, Vol. 34, No. 4, pp. 510–517, 1998

Printed in Great Britain
0959-8049/98 $19.00+0.00

PII: S0959-8049(97)10076-4

Original Paper

Validation of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (QLQ-C30) as a Measure of Psychosocial Function in Breast Cancer Patients

S.-A. McLachlan,[1] G.M. Devins[2] and P.J. Goodwin[3]

[1]Department of Oncology, St Vincent's Hospital, 41 Victoria Parade, Fitzroy 3065, Victoria, Australia; [2]University of Toronto, Clarke Institute of Psychiatry, University of Toronto, The Toronto Hospital, 250 College St, Toronto, Ontario M5T 1R8; and [3]The Department of Medicine, The Marvelle Koffler Breast Centre and the Samuel Lunenfeld Research Department at Mt Sinai Hospital, and the Clinical Epidemiology and Health Care Program, University of Toronto, 600 University Ave, Toronto, Ontario, M5G 1X5, Canada

The European Organisation for Research and Treatment of Cancer (EORTC) quality of life questionnaire QLQ-C30 has been developed as a quantitative measure of health-related quality of life for use in clinical trials of cancer patients. Validity is an important measurement property of all scientific tests. This study contributes to the iterative process of validating the questionnaire by focusing on the psychosocial subscales of the QLQ-C30, using baseline data from 150 patients participating in a randomised trial of supportive group therapy for metastatic breast cancer. The results provide strong support for the discriminative validity of the global health/quality of life, role function and social function subscales of the QLQ-C30, in patients differing according to clinical criteria. The psychosocial focus of the trial enabled expansion of criteria used to form patient subgroups, beyond medical factors, and consequently support was demonstrated for the discriminative validity of the emotional and cognitive function subscales. The degree of support for these subscales was less substantial than for the other QLQ-C30 subscales as there were fewer relevant criteria. Convergence assessed by correlations with independent measures of psychosocial function provides strong support for the convergent validity of the emotional function, role function and global health/quality of life subscales of the QLQ-C30, and moderate support for the social function subscales. There was little opportunity for the cognitive function subscale to associate with conceptually analagous subscales. Further testing is recommended with more comprehensive and specific measures of cognitive status. In general, the psychosocial subscales of the QLQ-C30 appear to be measuring the concepts they are purported to measure. © **1998 Elsevier Science Ltd. All rights reserved.**

Key words: quality of life, breast neoplasms, questionnaires, health status
Eur J Cancer, Vol. 34, No. 4, pp. 510–517, 1998

INTRODUCTION

BREAST CANCER is the second most frequent cause of cancer death among North American women [1, 2]. Metastatic breast cancer is incurable with currently available therapies, so the aim of treatment is to palliate symptoms, improve survival, and the quality of it. The benefits of cancer treatment should outweigh its cost in patient suffering.

Traditionally, treatment efficacy in breast cancer patients has been assessed in clinical trials using biomedical outcomes, such as tumour shrinkage, progression-free survival and treatment toxicity. These parameters may or may not correlate with patient benefit. Recently, it has been recognised that a more comprehensive assessment of the cancer patient is necessary and that the evaluation of outcomes must

Correspondence to S.-A. McLachlan.
Received 10 Jun. 1997; revised 22 Sep. 1997; accepted 17 Oct. 1997.

move beyond traditional biomedical endpoints to include assessments of the impact of disease and its treatment on patients' quality of life.

The European Organization for Research and Treatment of Cancer (EORTC) has developed a 30-item quality of life questionnaire (QLQ-C30) which is a brief, self-reporting, cancer-specific measure of health-related quality of life (HRQL) [3]. Its purpose is to obtain information about the impact of disease and treatment on the daily living of cancer patients. The core, QLQ-C30, is composed of a number of subscales representing HRQL dimensions relevant across a wide range of cancer sites and treatment methods. It can be supplemented by diagnosis-specific and treatment-specific questionnaire modules [4].

To be useful in research and clinical applications, HRQL measures, like other scientific measures, must be relevant, quantifiable, reliable and valid. There is a growing literature supporting the reliability and validity of the QLQ-C30 [3, 5–10]. These studies have focused on disease and treatment parameters such as performance status, chemotherapy, radiation treatment and stage of disease. The psychometric properties of the psychosocial subscales of the QLQ-C30 have not been well established. However, the QLQ-C30 is a multidimensional measure of HRQL and the focus of this study is on the items measured by the psychosocial subscales. We have defined the psychosocial subscales of the QLQ-C30 as those measuring role function, social function, cognitive function, emotional function and global/overall quality of life.

The reliability and validity of commonly employed psychological instruments is difficult to assess in cancer patients because somatic manifestations of psychological distress, such as fatigue, insomnia and anorexia, can also be attributed to disease or the treatment of it. As a compromise, selected items have been incorporated into multidimensional HRQL instruments to minimise respondent burden. The question is how well do abbreviated scales measure social and psychological functioning in cancer patients?

This study formally evaluates the construct validity of the psychosocial subscales of the QLQ-C30 in patients with metastatic breast cancer. The specific objectives were to examine the discriminative validity of the five psychosocial subscales and to evaluate their convergent validity.

PATIENTS AND METHODS

Subjects

The data are based on the first 150 women participating in a Canadian multicentre, randomised trial designed to determine the effectiveness of Breast Expressive Supportive Therapy (BEST) in metastatic breast cancer [11]. Women were eligible to enter the BEST study if they had metastatic cancer beyond the breast and ipsilateral axilla, an expected survival of at least 3 months and the absence of serious psychiatric disorders (psychosis, untreated major depression, severe personality disorder). Patients were required to speak and read English sufficiently well to complete psychological questionnaires and participate in a support group. All patients provided written informed consent to participate in the study.

Study measures and procedures

EORTC QLQ-C30. The EORTC QLQ-C30 is a 30-item questionnaire composed of five multi-item functional subscales: physical, role, emotional, social and cognitive functioning; three multi-item symptom scales measure fatigue, pain and emesis; global health/quality of life subscale; and six single items to assess financial impact and symptoms such as dyspnoea, sleep disturbance, appetite, diarrhoea and constipation [3]. The QLQ-C30 was scored according to algorithms recommended by the EORTC [12]. Higher scores represent better function or higher levels of symptoms. The conceptual and methodological issues underlying the development of the instrument are described elsewhere [12].

Additional study measures. Patients completed a battery of self-reporting questionnaires at baseline and those relevant to this validation study include the Psychosocial Adjustment to Illness Scale (PAIS) which is comprised of 46 items designed to assess a patient's psychosocial adjustment to conditions ranging from mild disorders to life threatening diseases [13]; the Profile of Mood States (POMS) is a 65-item adjectival check list which is intended to assess transient affective states [14]; the Mental Adjustment to Cancer scale (MAC) is a 40-item questionnaire assessing the extent to which patients adopt certain responses in their adjustment to the diagnosis and treatment of cancer [15]; the Impact of Event Scale (IES) is a 15-item scale developed to assess the subjective impact of any stressful life event [16], in this study it refers to actual recurrence of breast cancer; two 10 cm vertical linear analogue scales (LASA) measuring pain and suffering [17]; the Family Information Form which enquires about social environment and social support (D. Speigel, Stanford University, School of Medicine, California, U.S.A.). Raw scores from the POMS, PAIS, MAC, IES and LASA scales were used in the analysis. Generally, higher scores on these instruments represent worse mood and adjustment. 'Vigour' on the POMS and 'fighting spirit' on the MAC are negatively related to the other subscales. Clinical variables were recorded and included age, time with metastatic disease, ECOG performance status, current treatment and disease sites [18].

Data collection procedures. The aspects of validity considered in this study are cross-sectional in that they are based on a single administration of patient-based measures, prior to the start of the intervention. Sociodemographic and clinical data were also collected and reported at baseline.

Analysis plan. Non-parametric statistics are emphasised throughout the analysis. This decision was based on the ordinal nature of the patient-based measures and the non-normality of distributions, to limit assumptions and maintain consistency. A number of analyses were used to evaluate the validity of the five psychosocial subscales of the EORTC QLQ-C30. The first approach was applied to examine the discriminative validity of the instrument. Discriminative validity refers to the ability to detect cross-sectional differences between subjects when differences are expected (also termed known-groups comparison). External parameters available in the study were used to divide the patients into subgroups that we expected to differ in HRQL. These parameters were selected because they are clinically important, they were measured in the study independently of the QLQ-C30 and they might reasonably be expected to affect functioning and well being. They included clinical and psychosocial variables. Patients were divided on the basis of ECOG performance status (ECOG scores 0 versus 1–2), currently receiving chemotherapy (yes versus no), pain and suffering (lower quartile LASA score versus upper quartile LASA score), fatigue and sleep disturbance ($\leq$ median score versus $>$ median score), disease sites (non-visceral versus

visceral), living situation (alone versus with others), availability of a special person to communicate with (yes versus no), number of close relatives and friends (> 2 versus 0–2) and satisfaction with relationships (very–completely versus not very–somewhat). A priori hypotheses were made about the ability of the psychosocial subscales to discriminate between subgroups divided by these criteria. These predictions were made on the basis of the literature and a consensus approach between investigators (SAM, PJG). The Wilcoxon rank sum test was employed to test for the statistical significance of group differences [19]. With a fixed sample size, the lower the *P* value the greater the support for the discriminative ability. For descriptive purposes, the mean difference was employed to demonstrate the magnitude of HRQL differences between groups.

In the second approach, convergent validity of the instrument was evaluated. Convergent validity determines how closely the QLQ-C30 subscales relate to other independent measures of conceptually similar constructs. Divergent validity is supported when correlations with dissimilar traits or constructs are lower than with more closely related attributes. Bivariate relationships between the subscales of the QLQ-C30 and the other self-reporting instruments were examined with Spearman's rank correlation [18]. A priori predictions about the magnitude of inter-instrument correlations were made. A semi-formal judgemental approach was applied to define these hypotheses [20]. Two academic medical oncologists made and recorded their hypotheses independently. Predictions regarding the strength of inter-instrument correlations were based on an ordinal scale; negligible weak, moderate or strong. Results were reviewed and when discrepancy occurred between raters there was discussion and consensus was reached. Criteria for the quantitative significance of correlations were based loosely on the recommendations of Burnand and associates: <0.3 negligible; 0.3–0.45 moderate; 0.45–0.6 substantial; and > 0.6 high [21]. Agreement between correlations predicted by experts and those observed was evaluated using the quadratically weighted kappa statistic [22]. This approach was used for descriptive purposes, to be systematic and to avoid the possibility of rationalising observed correlations. The guidelines of Landis and Koch were used to evaluate quantitative significance of kappa: >0.75 excellent agreement beyond chance; <0.4 poor agreement beyond chance; and 0.4–0.75 fair to good agreement [23].

RESULTS

Patient sociodemographic and clinical characteristics

Patient characteristics are summarised in Table 1. On average, patients had metastatic breast cancer for 12 months. They spoke English well enough to complete questionnaires and participate in group therapy. The majority of patients had a physician-estimated prognosis of greater than 12 months. Forty-five per cent had visceral sites of disease and approximately half the women had an Eastern Cooperative Oncology Group (ECOG) performance status rating of zero. Most were being treated actively with anti-tumour therapy.

Descriptive statistics

Completion of baseline self-reporting questionnaires was an eligibility criterion for the BEST study, therefore baseline QLQ-C30 questionnaires were completed by and available for all 150 patients. The lowest response rate for individual subscales of the instrument was 147/150 (98%) for the emotional function, cognitive function and social function items. The lowest response rate for subscales of the comparison questionnaires was 131/150 (87%) for sexual relationships and 138/150 (92%) for domestic environment on the PAIS.

Table 1. *Baseline patient characteristics in the BEST study (*n = 150*)*

Age	Median (range)	49 years (29–79)	
Time with metastases	Median (range)	12 months (1–216)	
		n	(%)
Performance status	0	63	(42)
	1	64	(43)
	2	23	(15)
Current treatment	Chemotherapy	57	(38)
	Endocrine therapy	68	(45)
	Radiation	7	(5)
	No active	18	(12)
Disease sites	Visceral	67	(45)
	Non-visceral	83	(55)
Marital status	Married	105	(70)
	Divorced	15	(10)
	Widowed	7	(5)
	Other	23	(15)

All 5 psychosocial subscales of the QLQ-C30 had asymmetric distributions and were negatively skewed, indicating a predominance of higher scores or better functioning and HRQL. On average, patients had relatively high levels of cognitive function (Table 2). The median score for social function, emotional function and global health/quality of life was 67 (0–100). Patients had most trouble with role function. All the POMS and PAIS subscales, except vigour and health care orientation, were positively skewed, indicating lower levels of mood disturbance and maladjustment. Subscales on the MAC and IES have close to a normal distribution except MAC avoidance (data not shown).

Discriminative validity

The results of the assessment of discriminative validity are shown in Figures 1 and 2 which show the mean differences for each QLQ-C30 subscale between subgroups of patients expected to be different. The global health/quality of life, role function and social function subscales of the QLQ-C30 were able to detect all the anticipated differences in HRQL in subgroups defined by the clinical parameters of ECOG performance status, pain, suffering, fatigue and chemotherapy treatment. Differences were also detected in role and social function for patients divided according to suffering and chemotherapy. These were not predicted a priori. The emotional function subscale detected the anticipated differences

Table 2. *Descriptive statistics of QLQ-C30*

Psychosocial subscales	Items*	Median	LQ	UQ
Global quality of life	29, 30	67	50	83
Role function	6, 7	50	50	100
Social function	26, 27	67	50	100
Cognitive function	20, 25	83	67	100
Emotional function	22–24	67	58	83

*Numbers correspond to the item numbers in the questionnaire.
LQ, lower quartile, UQ, upper quartile.

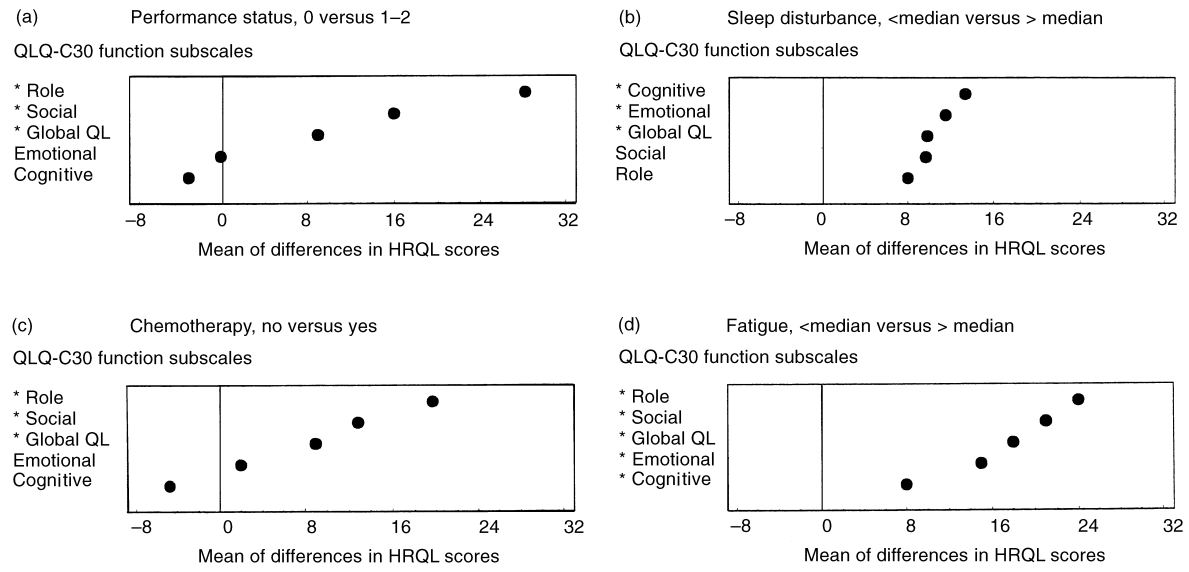

Figure 1. Differences in the scores of QLQ-C30 subscales in subgroups divided according to (a) performance status; (b) sleep disturbance; (c) receiving chemotherapy or not and; (d) fatigue. *Refers to a statistically significant difference (*P* value < 0.05). A positive score indicates a difference in the expected direction.

in subgroups defined according to suffering, fatigue and sleep disturbance but not in those with different pain scores. Cognitive function was different in subgroups defined by fatigue and sleep disturbance, but there was no support for its discriminative ability with reference to the other external clinical parameters, as anticipated. Site of metastatic disease (visceral versus non-visceral) did not distinguish between groups of patients differing in the clinical parameters (data not shown). Of the 19 hypothesised differences, there was supportive evidence for 17 and we failed to predict two significant differences (Table 3).

The magnitude of the differences in HRQL between subgroups according to clinical parameters was greatest for the role function, social function and global health/quality of life subscales, ranging from a mean difference of 9 to 28 (Figures 1 and 2). The results indicate that the global health/quality of life, role function and social function subscales have better cross-sectional discrimination when groups were divided according to conventional clinical measures, than the emotional and cognitive subscales. The statistical significance of the mean differences was tested using the unpaired *t*-test. The *P* values closely approximated those from the Wilcoxon rank sum test and no conclusions about the discriminative validity were affected.

The psychosocial subscales of the QLQ-C30 were not able to detect any of the anticipated differences in HRQL in

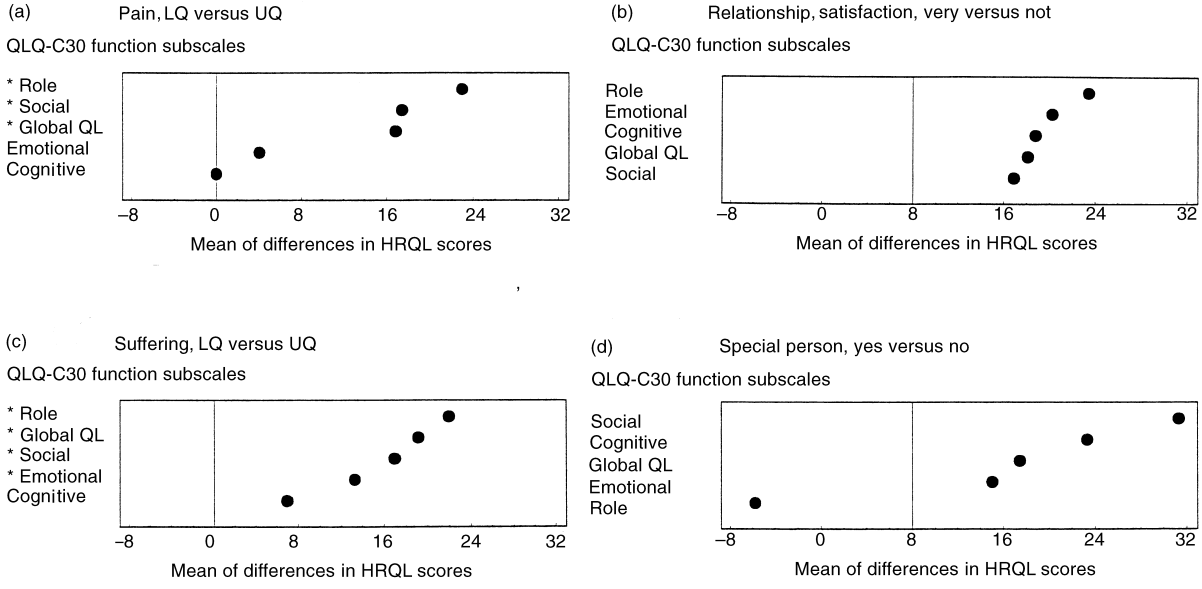

Figure 2. Differences in the scores of QLQ-C30 subscales in subgroups divided according to (a) pain; (b) satisfaction with relationships or not; (c) suffering; and (d) presence of special person. *Refers to a statistically significant difference (*P* value < 0.05). A positive score indicates a difference in the expected direction. LQ, lower quartile; UQ, upper quartile.

Table 3. *Agreement of hypothesised and observed differences in health-related quality of life (HRQL) between subgroups divided according to external parameters*

External criteria	QLQ-C30 subscales				
	Global quality of life	Role	Social	Emotional	Cognitive
Performance status	+	+	+		
Pain	+	+	+	–	
Suffering	+		+	+	
Fatigue	+	+	+	+	+
Chemotherapy	+	+	+		
Sites of disease		–			

+indicates that a difference in HRQL was anticipated between groups, and was substantiated. –indicates that a hypothesised difference was not substantiated. In some instances no hypotheses were made.

subgroups defined by the social support parameters of living situation, the presence of close personal contacts, or satisfaction with relationships. Although not statistically significant, there was a trend for better emotional function in those living with others and those satisfied with their relationships (Figure 2). The mean of the differences for emotional function in subgroups divided according to these criteria, seen in the dot plots, are 9 and 6, respectively. There was also a trend for social function to be better (mean difference of 12) in those who had a special person to communicate with (Figure 2).

The results of complementary analyses using Spearman's rank correlation coefficient, which quantifies the relationship between HRQL scores and criterion external variables, are shown in Table 4 and are consistent with the pattern observed by dichotomising and comparing two groups as described above.

Convergent validity

Tables 4–6 summarise bivariate associations between the five subscales of the QLQ-C30 and the four comparison instruments. Specific points will be illustrated with tables highlighting selected coefficients from these matrices. With a sample size of 150, correlation coefficients above 0.20 are statistically significant ($P < 0.01$).

A correlation matrix of the five psychosocial subscales of the QLQ-C30 and the PAIS is shown in Table 5. As expected, the strongest relationships occurred between conceptually analogous subscales, like the QLQ-C30 role function and the PAIS vocational ($r = 0.57$) and domestic environment ($r = 0.52$) subscales; the QLQ-C30 social function and the PAIS social ($r = 0.57$) and domestic ($r = 0.60$) environment subscales; and the QLQ-C30 emotional function and the PAIS psychological distress ($r = 0.68$) subscales. As hypothesised, the global health/overall quality of life subscale correlated moderately with related but distinct concepts measured by the PAIS, such as vocational ($r = 0.45$), domestic ($r = 0.54$) and social environment ($r = 0.54$) and psychological distress ($r = 0.39$) subscales.

Table 4. *Discriminative validity of the QLQ-C30 assessed by Spearman's rank correlations with independent criterion variables scored continuously. Higher correlations reflect better discrimination. A correlation of* 0.2 *is significant at the* 0.01 *level*

External parameter	Spearman's rank correlation with QLQ subscales				
	Global quality of life	Role	Social	Emotional	Cognitive
ECOG performance status	0.23	0.42	0.36	0.02	0.08
Pain	0.36	0.34	0.26	0.11	0.02
Suffering	0.42	0.27	0.28	0.25	0.17
Fatigue	0.56	0.37	0.49	0.44	0.32
Chemotherapy	0.22	0.27	0.26	0.08	0.10
Sleep disturbance	0.35	0.17	0.23	0.33	0.45

Table 5. *Convergent validity. Spearman's rank correlations between the QLQ-C30 psychosocial subscales and the Psychosocial Adjustment to Illness Scale (PAIS)*

PAIS subscales	Spearman's rank correlation with QLQ-C30 subscales				
	Global quality of life	Role	Social	Emotional	Cognitive
Health care orientation	0.24	0.21	0.16	0.35	0.25
Vocational environment	0.45†	0.57†	0.48	0.24	0.11
Domestic environment	0.54†	0.52*	0.60*	0.35	0.33
Sexual relationships	0.35	0.28	0.41†	0.22	0.09
Extended family	0.29	0.15	0.26*	0.27	0.30
Social environment	0.54†	0.39	0.57*	0.40	0.26
Psychological distress	0.39†	0.25	0.40	0.68*	0.29
PAIS total	0.63*	0.60†	0.66†	0.53†	0.32

*Correlations hypothesised a priori to be 'strong'. †Those hypothesised to be 'substantial'. Unmarked coefficients were those expected to be 'negligible or weak'.

Table 6. *Convergent validity. Spearman's rank correlations between the QLQ-C30 psychosocial subscales and the Profile of Mood States (POMS)*

	Spearman's rank correlation with QLQ-C30 subscales				
POMS subscales	Global quality of life	Role	Social	Emotional	Cognitive
Tension	0.48	0.24	0.37	0.76*	0.46
Depression/dejection	0.40	0.17	0.25	0.74*	0.29
Anger	0.28	0.08	0.19	0.55†	0.25
Vigor/activity	0.49	0.24	0.31	0.36†	0.31
Fatigue/inertia	0.56†	0.37	0.49†	0.44†	0.32
Confusion/bewilderment	0.35	0.07	0.21	0.59	0.54*
Total mood disorder	0.56†	0.27	0.38	0.74*	0.45

*Correlations hypothesised a priori to be 'strong'. †Those hypothesised to be 'substantial'. Unmarked coefficients were those expected to be 'negligible or weak'.

The PAIS total score is a summation of all subscales and therefore has complementary relationships with all the psychosocial subscales of the QLQ-C30, except cognitive function, which did not converge with any of the constructs measured by the PAIS. Health care orientation (PAIS) was not associated with any of the QLQ-C30 subscales, as was predicted and specified a priori.

Strong inter-instrument associations occurred between POMS and QLQ-C30 subscales measuring similar phenomena (Table 6). Tension, depression/dejection and the POMS total mood disturbance subscales correlated very strongly with emotional function ($r > 0.74$). This was expected a priori since the emotional function subscale of the QLQ-C30 focuses on anxiety and depression. Moderate correlations were also seen between anger, fatigue/inertia, confusion/bewilderment and the emotional subscale of the QLQ-C30. In a priori predictions, we underestimated the strength of association between confusion/bewilderment and QLQ-C30 emotional function. On reviewing the literature, it is clear that our hypothesis was incorrect. In previous studies the cognitive subscale correlated well with the emotional subscale of the Cancer Rehabilitation Evaluation System (CARES) and the overall score on the GHQ [10]. Intra-instrument correlations between the QLQ-C30 emotional function and cognitive function subscales have also been moderately strong. The correlation between confusion/bewilderment and the cognitive function items of the a QLQ-C30 was substantial although not as high as expected, as the two subscales were thought to be measuring corresponding concepts.

The global health/quality of life subscale was substantially correlated with the total mood disturbance score ($r = 0.56$) and other related but distinct states, particularly fatigue/inertia, tension and vigour/activity. Many of the associations between QLQ-C30 role function, social function, cognitive function and the POMS subscales were weak or absent, as expected.

In general, the inter-instrument associations between the MAC and QLQ-C30 were lower than for the POMS or PAIS (Table 7). As predicted a priori, there were weak to moderate bivariate correlations between QLQ-C30 emotional function and MAC fighting spirit, hopeless/helpless and anxious preoccupation. None of the subscales were thought to be measuring the same concepts, but, we predicted that social function would relate substantially to coping style, as measured by the MAC. This hypothesis was not substantiated.

The only quantitatively significant correlation occurred between emotional function and intrusion (IES). It was predicted that both intrusion and avoidance would be related to emotional function, as measured by the QLQ-C30 (Table 7).

Predictions based on all the comparison questionnaires were most accurate for the global health/quality of life, emotional function and role function subscales. The agreement

Table 7. *Convergent validity. Spearman's rank correlations between the QLQ-C30 psychosocial subscales and the Mental Adjustment to Cancer Scale (MAC) and Impact of Event Scale (IES)*

	Spearman's rank correlation with QLQ-C30 subscales				
	Global quality of life	Role	Social	Emotional	Cognitive
MAC subscales					
Fighting spirit	0.34†	0.17	0.12†	0.38*	0.23
Hopeless/helpless	0.29	0.20	0.18†	0.48†	0.15
Anxious preoccupation	0.32	0.15	0.15†	0.45†	0.18
Fatalism	0.12	0.06	0.02	0.16	0.12
Avoidance	0.08	—	0.03	0.12	0.01
IES subscales					
Intrusion	0.28	0.03	0.24*	0.55†	0.31
Avoidance	0.14	0.04	0.13*	0.31†	0.14

*Correlations hypothesised a priori to be 'strong'. †Those hypothesised to be 'substantial'. Unmarked coefficients were those expected to be 'negligible or weak'.

between hypothesised and observed correlations was excellent for these subscales. There was fair to good agreement between predicted and observed associations for the QLQ-C30 social function and cognitive function subscales.

Agreement between predicted and observed inter-instrument relationships was highest for the QLQ-C30 with the POMS (quadratically weighted Kappa, $Kw = 0.74$) and the QLQ-C30 with the PAIS ($Kw = 0.65$), but moderate for the QLQ-C30 with the IES ($Kw = 0.63$), and lowest for the QLQ-C30 with the MAC ($Kw = 0.47$).

In view of the inconsistencies identified in previous psychometric evaluations of the MAC, our predictions were tentative. The difficulty is reflected in the low concordance between expected and observed correlations for the QLQ-C30 and the MAC. Excluding the MAC data altered the degree of concordance between observed and expected correlations only for the social function subscale, in that the kappa value was higher ($Kw = 0.60$) which suggests more support for the convergent validity of the social function subscale.

DISCUSSION

The QLQ-C30 is a measure of HRQL, intended to be applicable across a wide range of cancer diagnoses and treatments. Breast cancer is a common malignancy and women with the disease represent a large portion of oncology patients. The QLQ-C30 has been evaluated in other diagnostic groups of patients and smaller subgroups of patients with heterogeneous cancers, including breast cancer. The results presented in this paper represent one of the largest validation studies involving metastatic breast cancer patients. They contribute to the iterative process of validating the QLQ-C30 as a multidimensional HRQL measure, applicable across a variety of cancer types. The findings support the cross-sectional validity of the QLQ-C30 subscales for global health/quality of life, role function, social function and emotional function. There is sufficient evidence in support of the validity of the cognitive function subscale to encourage further evaluation in future studies. The psychosocial subscales of the QLQ-C30 appear to be measuring what they are intended to measure.

The global health/quality of life subscale of the QLQ-C30 places emphasis on the physical aspects of health by including two items intended to measure overall physical condition and quality of life. It therefore discriminated well when groups were divided according to disease and treatment status. It also correlated with distinct but related physical, social and emotional concepts. This suggests that a variety of components contribute to patients' perception of their overall quality of life. A separate global subscale is preferred as it allows individual differences in the evaluation of information people have about their disease. It also extends measurement into the well-being range, not solely focusing on negative aspects of HRQL.

The role function subscale is comprised of two items addressing the effect of disease on the ability to carry out job and home duties. It discriminated well on the basis of physical, clinical criteria. The role function items of the QLQ-C30 did not distinguish between social support parameters, as measured by the Spiegel social support questionnaire. In most populations, limitations in role performance are due to physical health problems [24]. The role function subscale associated strongly with other measures of similar concepts and moderately with related but distinct attributes. There is good support that the role function subscale is measuring what it is intended to assess. The dichotomous response format is probably too coarse and restricts the range of possible scores. The scale has been modified by the EORTC quality of life study group, to a four point Likert type and is currently undergoing testing.

The effect of disease on the ability to interact with family and friends is measured by the two-item social function subscale. There is considerable support for the convergent validity of the social function subscale. It converged particularly well with PAIS subscales measuring similar attributes. Correlations between distinct but related concepts were not as strong as anticipated. The social function items, like the role items, are physically weighted and relate particularly to physical and functional aspects of these dimensions rather than to emotional and affective components, as measured by most of the comparison questionnaires. This may account for the relative weakness of distinct but related relationships. There is good support for the discriminative validity of the social function subscale for subgroups divided according to clinical parameters. Contrary to expectations, the data do not provide evidence for the ability of the subscale to differentiate on the basis of social support. There is no published information regarding the measurement properties of the social support questionnaire used in this study. Important components of social support and adjustment include the degree of social integration of the patient in the family, at the work place, and in the larger community, as well as satisfaction with major social roles and interpersonal ties. It is important to recognise that difficulties exist in obtaining data reflecting the actual social reality of patients. Problems of social desirability in response patterns can occur, particularly when assessing the quality of relationships. This effect may have contributed to the skewed distributions of social support variables seen for women in the BEST study.

The emotional function subscale is comprised of four items that focus on anxious and depressed effect. It associated very strongly with analogous items from the comparison instruments. This finding adds strong supportive evidence to the existing data regarding the validity of the emotional subscale. As expected, and as reported by other investigators, subgroups differing by conventional clinical standards did not differ with regard to emotional function. By employing a broader range of discriminating criteria and ones that were more likely to be related to the emotional and cognitive function of patients, this study provides the first evidence of support for the discriminative validity of the emotional and cognitive function subscales. For example, subgroups divided according to fatigue had detectable differences in emotional and cognitive function. To be more confident about the discriminative validity of these subscales, further investigation is required. Expansion of the scope and relevance of discriminating criteria should be pursued in future studies.

Cognitive function is represented by two items evaluating concentration and memory. Patients participating in the BEST study (and other clinical trials evaluating HRQL) were required to have unimpaired cognitive function to be able to participate in group therapy and to complete the battery of patient-based measures. Consequently the distribution of scores was highly skewed. The restricted range of cognitive function scores is likely to have affected the ability to distinguish between different subgroups and to have affected associations with other patient-based questionnaires. None the less, the

data from this analysis demonstrate good support for the cognitive items. It was able to detect discrepancies in sleep disturbance and fatigue, and it converged moderately with the single subscale on the POMS measuring an analogous concept. More data are required before the cognitive function subscale can be used with confidence.

The BEST trial was not designed specifically to evaluate the validity of the QLQ-C30. However, examination of the five psychosocial subscales was facilitated in the context of a psychosocial intervention trial. These dimensions were also studied in more detail because other groups evaluating the QLQ-C30 have focused on conventional clinical parameters. Limitations occurred because the analysis was confined to data that were collected and available in the BEST study, and to a cross-sectional analysis because the trial is ongoing. A major disadvantage was the inability to analyse longitudinal data or to make comparisons across groups. This would have broadened assessment of discriminative validity and may have allowed an evaluation of predictive validity. Doing so, however, would have jeopardised the integrity of the final trial results and the feasibility of ongoing randomisation, if a difference was identified between groups. Specific entry criteria for the BEST study ensured excellent compliance with patient-based measures. However, they also imposed restrictions on the sample of patients. The distributions of the patient-based measures were skewed towards better HRQL and psychosocial function. This made analysis and interpretation of the data challenging and may have made it more difficult to detect true differences in the five subscales than would have emerged with a more heterogeneous population.

When treating cancer patients, preservation and/or improvement of quality of life is an important goal. There is an emerging consensus that a pragmatic definition of quality of life be adopted and operationalised as an outcome in clinical trials. HRQL is a functional representation of patients' physical, psychological, and social response to disease and its treatment. The QLQ-C30 has been developed as a quantitative measure of HRQL for the clinical research setting. As is the case for other scientific tests, application is dependent on the instrument's consistency and accuracy. Validation is relative and its evaluation is an iterative process. Tests of validity do not provide absolute proof or rejection but rather a continuum of support and are best applied in the context in which the instrument is intended to be used, as was the case in this analysis.

The results from this study contribute to the growing literature supporting the scientific validity of the QLQ-C30 as a cancer-specific measure of HRQL. The data collected in the BEST study were novel in comparison to conventional clinical parameters, commonly the focus of cancer clinical trials and validation studies of HRQL instruments. Included were a battery of established psychological questionnaires and information regarding social circumstances. The contribution of the research is unique in that support is specifically shown for the psychosocial subscales of the QLQ-C30.

1. Boring CC, Squires TS, Tong T, *et al.* Cancer statistics, 1994. *CA Cancer J Clin* 1995, **44**, 7–26.
2. National Cancer Institute of Canada. *Canadian Cancer Statistics 1996*. Toronto, Canada, 1996.
3. Aaronson NK, Ahmedzai S, Bergman 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.
4. Aaronson NK, Bullinger M, Ahmedzai S. A modular approach to quality-of-life assessment in clinical trials. *Recent Results Cancer Res* 1988, **111**, 231–245.
5. Bjordal K, Kaasa S. Psychometric validation of the EORTC core quality of life questionnaire, 30-item version and a diagnosis-specific module for head and neck cancer patients. *Acta Oncol* 1992, **31**, 311–321.
6. Osoba D, Zee B, Warr D, Kaiser 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.
7. Ringdal GI, Ringdal K. Testing the EORTC quality of life questionnaire on cancer patients with heterogeneous diagnoses. *Qual Life Res* 1993, **2**, 129–140.
8. Jensen Hjermstad M, 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**, 1249–1253.
9. Bjordal K, Kaasa S. Psychometric validation of the EORTC core quality of life questionnaire, 30-item version and a diagnosis-specific module for head and neck cancer patients. *Acta Oncol* 1992, **31**, 311–321.
10. Niezgoda HE, Pater JL. A validation study of the domains of the core EORTC Quality of Life Questionnaire. *Qual Life Res* 1993, **2**, 319–325.
11. Goodwin PJ, Leszcz M, Koopmans J, *et al.* Randomized trial of group psychosocial support in metastatic breast cancer: the BEST study. *Cancer Treatment Rev* 1996, **22**, 91–96.
12. Fayers P, Aaronson NK, Bjordal K, Sullivan M for the EORTC Quality of Life Study Group. *EORTC QLQ-C30 Scoring manual*, 1995.
13. Derogatis LR, Derogatis MF. *The Psychosocial Adjustment to Illness Scale (PAIS and PAIS-SR): Scoring, Procedures and Administration manual-II*. Baltimore, Clinical Psychometric Research, 1990.
14. McNair DM, Lorr M, Droppleman LF. *EdITS Manual for the Profile of Mood States*. San Diego, EdITS/Educational and Industrial Testing Service, 1992.
15. Watson M, Greer S, Young J, *et al.* Development of a questionnaire measure of adjustment to cancer: the MAC scale. *Psychol Med* 1988, **18**, 203–209.
16. Horowitz MJ, Wilner N, Alvarez W. Impact of Event Scale: a measure of subjective stress. *Psychosom Med* 1979, **41**, 209–218.
17. Spiegel D, Bloom JR. Pain in metastatic breast cancer. *Cancer* 1983, **52**, 341–345.
18. Zubrod CG, Scheiderman M, Frei E, *et al.* Cancer—appraisal of methods for the study of chemotherapy of cancer in man: thiophosphamide. *J Chron Dis* 1960, **11**, 7–33.
19. Lehmann EL. *Non-Parametrics: Statistical Methods Based on Ranks*. San Francisco, Holden-Day, 1975.
20. Fink A, Kosecoff J, Chassin M, *et al.* Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984, **74**, 979–983.
21. Burnand B, Kernan WN, Feinstein AR. Indexes and boundaries for 'quantitative significance' in statistical decisions. *J Clin Epidemiol* 1990, **43**, 1273–1284.
22. Kramer MS, Feinstein AR. Clinical biostatistics. The biostatistics of concordance. *Clin Pharmacol Ther* 1981, **29**, 111–123.
23. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977, **33**, 159–174.
24. Ware JE. *SF-36 Health Survey Manual and Interpretation Guide*. Boston, Nimrod Press, 1993.

Acknowledgements—We would like to acknowledge the contributions of the following BEST study co-investigators: Coordinating Centre, Molyn Leszcz MD, John Hunter MD, Leslie Vincent RN, Kathy Pritchard MD; Ottawa Centre, Margaret Navarro MD, Paul Aucoin MSW, Shail Verma MD; Hamilton Centre, Julia Masterson MD, Helaine Guther MSW, Andrew Arnold MD; Winnipeg Centre, Harvey M, Chochinov MD, Barbara Warren RN; Edmonton Centre, Marilyn Hundleby PhD, Rami Sela PhD, Jean-Marc Nabholtz MD; Calgary Centre, Jan Koopmans MSW, Michael Speca PhD, Alexander Patterson MD; Vancouver Centre, Elaine Drysdale MD, Elizabeth Dohan MSW, Richard Doll PhD, Susan O'Reilly MD.