Pergamon

*European Journal of Cancer*, Vol. 34, No. 6, pp. 810–815, 1998

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

PII: S0959-8049(97)00395-X

## Original Paper

# A Cross-validation of the European Organization for Research and Treatment of Cancer QLQ-C30 (EORTC QLQ-C30) for Japanese with Lung Cancer

K. Kobayashi,[1,2] F. Takeda,[1] S. Teramukai,[3] I. Gotoh,[1] H. Sakai,[1] S. Yoneda,[1] Y. Noguchi,[1] H. Ogasawara[1] and K. Yoshida[1]

[1]Hospital of Saitama Cancer Centre, Saitama; [2]Fourth Department of Internal Medicine, Nippon Medical School, 1-1-5 Sendagi, Bunkyo-ku; and [3]School of Health Sciences, University of Tokyo, Tokyo, Japan

**The EORTC QLQ-C30 was developed in English-speaking cultures. To determine if this instrument could cross a broad cultural divide and be used in Japan, the cross-cultural validity of its Japanese version was estimated. In evaluating psychometric testing, internal consistency by Cronbach's alpha, item-discrimination by multitrait scaling analysis, and validity analysis with ECOG performance score (PS) and Karnofsky Performance Status Scale (KPS) were performed. The QLQ-C30 (version 1.0) was given to 105 patients with lung cancer. Although the response rate was low in patients with PS 4, the questionnaire was well accepted by patients with PS 0–3. The Japanese QLQ-C30 has a weak scale of role functioning in terms of item discriminative validity. It also has a weak scale of cognitive functioning in items of discriminative validity and internal consistency. However, known-groups comparison showed the expected clinical validity with PS for all the scales except for financial impact, and longitudinally clinical validity with KPS was shown in scales of cognitive functioning, fatigue, and nausea and vomiting. Multitrait scaling analysis showed that the predicted scales constituting quality of life (QOL) in the English-speaking culture were extracted from the Japanese QLQ-C30, and found to be valid in Japan, indicating its possible usefulness as an instrument that is universally applicable across cultures.** © **1998 Elsevier Science Ltd. All rights reserved.**

**Key words: quality of life, EORTC QLQ-C30, feasibility, performance status, cross-cultural comparison, Japanese, lung carcinoma**
*Eur J Cancer*, Vol. 34, No. 6, pp. 810–815, 1998

## INTRODUCTION

MOST PATIENT-CENTRED questionnaires on quality of life (QOL) have been developed in English-speaking cultures. Thus, it has been questioned as to whether or not the QOL questionnaires could cross a broad cultural divide and be used in Eastern countries.

The European Organization for Research and Treatment of Cancer (EORTC) developed a QOL questionnaire, the EORTC Core Quality of Life Questionnaire (EORTC QLQ-C30, version 1.0) [1]. Its validity has already been demonstrated [1–4]. Up to now, the English version has been translated into more than 25 languages. A Japanese version of the EORTC QLQ-C30 was also drawn up by the EORTC itself. Aaronson and associates examined the differences in the scores according to language and culture in European countries [1], but the comparison of results in Japan have not been reported because of the limited numbers of Japanese cases examined [1]. In fact, the validity of the EORTC QLQ-C30 has not been reported in any country in the Far East, and there is a question of whether the EORTC QLQ-C30 is a universally applicable instrument in Eastern cultures.

Though there is still no standard method ('gold standard') for cross-cultural validation [5], Hui and Triandis proposed a dimension for cross-cultural validation [5, 6]. They postulated four dimensions of equivalence when attempting to internationally measure a construct such as QOL. These dimensions are functional equivalence (adequacy of translation),

Correspondence to K. Kobayashi.
Received 4 Apr. 1997; revised 28 Jul. 1997; accepted 1 Aug. 1997.

scale equivalence (comparability of response scales), operational equivalence (standardization of psychometric testing procedures) and metric equivalence (transferability of scoring results from one culture to another) [5, 6].

In this study, we examined the cross-cultural validation in the possible dimensions proposed by Hui and Triandis using the Japanese EORTC QLQ-C30 described above.

## PATIENTS AND METHODS

*Study sample and protocol*

Subjects were patients with lung cancer who were admitted to the pulmonary section of the Saitama Cancer Center from March to September 1992. After informed consent had been obtained to estimate their QOL, the Japanese EORTC QLQ-C30 was administered on Wednesdays every 2 weeks during each patient's hospitalisation. At the same time, their doctors were asked questions concerning ECOG performance score (PS), the Karnofsky Performance Status Scale (KPS) [7] and the treatment every 2 weeks.

*Instruments used in the study*

The Japanese EORTC Core Quality of Life Questionnaire was developed by EORTC itself in cooperation with the Saitama Cancer Center from the original QLQ-C30 (version 1.0) [1]. A rigorous translation/back-translation process was performed [8]. This was followed by pilot testing with Japanese patients with lung cancer in the Saitama Cancer Center. In this pilot study, it was noticed that the polarity of answering was quite different between the English and Japanese versions in those questions which contained the term, 'any trouble', and appropriate modifications of the questions were made.

From the scales of the English EORTC QLQ-C30 reported [1], the scales of the Japanese version were assumed to have nine multi-item scales: five functional scales (physical, role, cognitive, emotional and social), a global QOL scale and three symptom scales (fatigue, pain and nausea/vomiting). It was also assumed to have six single-item symptom measures (dyspnoea, sleep disturbance, appetite loss, constipation, diarrhoea, financial impact). As in the English version, the seven items for physical functioning and role functioning employ a yes/no response choice, while the two items for the global QOL scale use a modified seven-point linear analogue scale. The other items employ Likert scale scoring [1].

*Psychometric testing for cross-cultural validation*

The return rate (feasibility) of the Japanese EORTC QLQ-C30 at an initial measure was calculated, and the feasibility for longitudinal measurements was also investigated.

After all the scores for the assumed scales/single items were linearly transformed to a 0–100 scale, their mean and standard deviation (S.D.) were calculated, and a frequency analysis was performed. A higher score represents a higher level of functioning on the functional scale, and a higher score represents a worse level of symptomology in measures of this domain.

A multitrait scaling analysis [9, 10] was carried out to evaluate the hypothesised scale structure of the questionnaire. This technique, to test for item convergence and discriminative validity, is based on the examination of item-scale correlations. Pearson's correlations of an item with its own scale (corrected for overlap) and other scales were calculated. Evidence of item convergence validity was defined as a correlation above 0.40 with its own scale. Item discriminant validity was supported by a comparison of the magnitude of the correlation of an item with its own scale compared with other scales. A definite scaling error was assumed if the correlation of an item with another scale exceeded the correlation with its own scale.

The internal consistency (reliability) of each scale was estimated by Cronbach's alpha coefficient [11]. A value of 0.70 or greater was considered acceptable for group comparison.

Construct validity was estimated by the correlations among the nine multi-item scales comprising the EORTC QLQ-C30. It was hypothesised that if there was a Pearson's correlation of more than 0.40 between scales, they were conceptually related. The weights for the global QOL scale of the other scales were also estimated by calculating the multiple regression equation when the global QOL scale was the criterion variable and the scores of the other scales were explanatory variables.

To evaluate the clinical validity of the Japanese QLQ-C30, the known-group comparison was performed [1]. This could indicate the extent to which the questionnaire scores were able to discriminate between subgroups of patients differing in clinical status. The clinical parameter, ECOG score, was employed to form mutually exclusive patient subgroups. Student's *t*-test was employed to test for the statistical significance of group differences. Finally, to determine longitudinally clinical validity, changes in the Japanese QLQ-C30 scores over time were examined in relationship to change in KPS. Respective improvement and deterioration was defined as a shift of at least one level upward and downward on KPS. Repeated-measures ANOVA was used to test for between-group differences over time in scores on the Japanese QLQ-C30.

## RESULTS

*Feasibility of the EORTC QLQ-C30 for Japanese*

105 patients with primary lung cancer were studied, and their characteristics are shown in Table 1. Most had progressive disease (stage $III_B$, 29 patients; stage IV, 68 patients). 8 patients had stage I–$III_A$ disease, but had small cell lung cancer or inoperable non-small cell lung cancer due to poor lung function. All were considered to have a poor prognosis.

Within 2 weeks before the initial measure of the EORTC QLQ-C30, 22 patients received chemotherapy containing cisplatin, 7 received chemotherapy without cisplatin, 11 received irradiation, and 52 received no active treatment (data missing for 13 patients). In the initial assessment, 12 patients did not answer the questionnaire: 1 refused to answer, while the other 11 patients could not answer because of poor performance status (PS 4); as a result feasibility was 89% (93/105). In subsequent assessments, an additional 11 of the 93 patients could not answer the QLQ-C30. Because of the study setting, in which the hospitalised patients in one ward were directly given the QLQ-C30 by his/her doctor, there was no administrative failure. The 11 patients were also estimated as PS 4 by their doctors at the time when they could not complete the QLQ-C30.

A total of 444 questionnaires were administered to these 105 patients (average 4.2/patient); 370 were completed, an overall response rate of 83%. In 344 assessments in which KPS was checked, although the response rate for PS 0–2

*Table* 1. *Characteristics of study sample* (n = 105)

| Characteristic | No. of patients (%) |
|---|---|
| Sex | |
| Male | 82 (78.1) |
| Female | 23 (21.9) |
| Histology | |
| SCLC | 25 (23.8) |
| NSCLC | 80 (76.2) |
| Staging | |
| I | 2 (1.9) |
| II | 3 (2.9) |
| $III_A$ | 3 (2.9) |
| $III_B$ | 29 (27.6) |
| IV | 68 (64.8) |
| ECOG PS | |
| 0 | 11 (10.5) |
| 1 | 37 (35.2) |
| 2 | 24 (22.9) |
| 3 | 13 (12.4) |
| 4 | 20 (19.0) |

SCLC, small cell lung cancer. NSCLC, non-small cell lung cancer. PS, performance status.

patients was over 99% (225/228), the corresponding rates for PS 3 and PS 4 were 81% (38/47) and 13% (9/69), respectively. Therefore, the questionnaire was well accepted in patients with a good performance status, but feasibility was low in patients with a poor performance status.

*Psychometric testing for cross-cultural validation*

For frequency analysis, mean scores ± S.D. of the assumed scales/single items on the questionnaire are shown in Table 2.

Multitrait scaling analysis of the QLQ-C30 showed that all item-scale correlations were above 0.40, indicating satisfactory item convergent validity (Table 3). In terms of item discriminant validity, one definitive scaling error in the role functioning scale was noted, and three definitive scaling errors were found in the cognitive functioning scale. The rate of scaling error was 2.1% (4/192). The low number of scaling errors supports the hypothesised scale structure of the Japanese QLQ-C30.

*Table* 3. *Multitrait scaling analysis for the EORTC QLQ-C*30

| | Number of scaling errors | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | PF | RF | CF | EF | SF | QL | FA | PA | NV |
| Japanese EORTC QLQ-C30 | | | | | | | | | |
| item convergent validity | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| item discriminative validity | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| Data reported by Aaronson and associates [1] | | | | | | | | | |
| item convergent validity | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| item discriminative validity | — | — | — | — | — | — | — | — | — |
| Data reported by Kaasa and associates [9] | | | | | | | | | |
| item convergent validity | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| item discriminative validity | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

PF, physical functioning; RF, role functioning; CF, cognitive functioning; EF, emotional functioning; SF, social functioning; QL, a global quality of life; FA, fatigue; PA, pain; NV, nausea/vomiting.

Cronbach's alpha coefficients for eight multi-item scales, except for cognitive functioning, was 0.70 or greater (Table 4), indicating satisfactory internal consistency. The lowest reliability was obtained for the cognitive functioning scale (0.63).

In interscale correlations (Table 5), the highest correlations were observed between the physical and role functioning scales ($r = 0.83$), as expected. Cognitive functioning was highly correlated with fatigue ($r = 0.70$) and emotional functioning ($r = 0.78$). Also, fatigue was highly correlated with emotional functioning ($r = 0.70$).

From a univariate analysis, global QOL correlated with physical functioning, role functioning, cognitive functioning, emotional functioning, fatigue, pain, nausea/vomiting sleep

*Table* 2. *Frequency analysis for EORTC Core Quality of Life Questionnaire (QLQ-C*30*)*

| | | for Japanese ($n = 93$) | | for Europeans ($n = 305$)§ | |
|---|---|---|---|---|---|
| | Items* | Mean score | S.D. | Mean score | S.D. |
| Functioning scales† | | | | | |
| Physical | 1–5 | 48.6 | 36.5 | 65.8 | 27.1 |
| Role | 6, 7 | 51.1 | 43.6 | 57.3 | 38.6 |
| Cognitive | 20, 25 | 64.7 | 28.0 | 83.6 | 20.5 |
| Emotional | 21–24 | 71.5 | 27.8 | 70.0 | 22.3 |
| Social | 26, 27 | 65.4 | 27.8 | 77.3 | 27.6 |
| Global quality of life | 29, 30 | 53.5 | 24.8 | 56.7 | 23.5 |
| Symptom scales and/or items‡ | | | | | |
| Fatigue | 10, 12, 18 | 46.6 | 29.8 | 39.4 | 25.4 |
| Pain | 9, 19 | 35.1 | 32.7 | 29.3 | 30.8 |
| Nausea and vomiting | 14, 15 | 16.6 | 24.9 | 6.7 | 15.5 |
| Dyspnoea | 8 | 30.5 | 31.0 | 41.0 | 28.7 |
| Sleep disturbance | 11 | 31.9 | 32.2 | 31.9 | 33.1 |
| Appetite loss | 13 | 33.7 | 34.6 | 26.9 | 35.1 |
| Constipation | 16 | 26.2 | 34.4 | 20.1 | 31.0 |
| Diarrhoea | 17 | 8.7 | 20.0 | 4.2 | 14.0 |
| Financial impact | 28 | 23.7 | 27.9 | 12.4 | 24.8 |

*Numbers correspond to item numbers on the QLQ-C30 (version 1.0). †Scores range from 0 to 100, with a higher score representing a higher level of functioning. ‡Scores range from 0 to 100, with a higher score representing a greater degree of symptoms. §Data reported by Aaronson and associates [1].

disturbance and appetite loss at more than 0.40 Pearson's correlations (Table 5). Multivariate analysis showed that emotional functioning and fatigue were significant explanatory variables for the global QOL scale (regression coefficient = 0.38, $P < 0.001$ and regression coefficient = 0.42, $P < 0.001$, respectively).

In the initial measure of the QLQ-C30 for each patient, expected clinical validity with ECOG PS was found for all the scales except for financial impact (Table 6). In the subsequent estimation, 22 patients had improved Karnofsky performance status scale (KPS) at more than one point, 24 patients had deteriorated KPS at more than one point and 22 patients had unchanged KPS. Repeated-measure ANOVA was employed to test for between-group (the three performance status subgroups) differences over time in the QLQ-C30 scores. Statistically significant between-group differences over time were observed for the scales of cognitive functioning, fatigue and nausea and vomiting (Table 7).

## DISCUSSION

Both the Japanese language and culture are vastly different from those of English speaking countries. Thus, in research of cross-cultural validation, it is of great interest whether instruments universally applicable can be validated cross-culturally in Japan. The main focus of this study was the cross-cultural validation of the EORTC QLQ-C30 in Japan. The four dimensions proposed by Hui and Triandis (i.e., functional, scale, operational and metric equivalence) require two steps in the approach to national application of internationally available instruments [6]. The two steps involve the creation (translation of items and responses) and evaluation (psychometric testing and norming) of an instrument [5].

The Japanese EORTC QLQ-C30 was created through 'the parallel approach' [5] by EORTC as previously reported [1, 8]. Repeated translation/back-translation processes on multiple occasions were performed. In this step, we had the problem which was lack of the term 'Quality of Life' in the Japanese language, and it was translated qualitatively. After achieving a consensus across paired samplings, this was followed by pilot testing with Japanese patients, which resulted in the Japanese EORTC QLQ-C30 with content and face validity.

Following the principle for evaluating psychometric testing, frequency analysis (Table 2), reliability analysis (Table 4), item-discrimination (Table 3) and validity analysis (Tables 5–7) of the Japanese EORTC QLQ-C30 were compared with Western results reported in the literature. In the frequency analysis, the scores on the Japanese EORTC QLQ-C30 were somewhat lower in functioning scales (i.e., poorer functioning) and higher in symptom scales and/or items (i.e., greater degree of symptoms) compared with the results of

*Table 4. Internal consistency (Cronbach's alpha) of the EORTC QLQ-C30*

| | Sample (*n*) | PF | RF | CF | EF | SF | QL | FA | PA | NV |
|---|---|---|---|---|---|---|---|---|---|---|
| Japanese QLQ-C30 | Lung cancer (93) | 0.83 | 0.74 | 0.63 | 0.90 | 0.72 | 0.88 | 0.90 | 0.87 | 0.86 |
| Data reported by | | | | | | | | | | |
| Aaronson and associates [1] | Lung cancer (305) | 0.68 | 0.54 | 0.56 | 0.73 | 0.68 | 0.86 | 0.80 | 0.82 | 0.65 |
| Bjordal and associates [2] | Head and neck cancer (126) | 0.74 | 0.74 | 0.28 | 0.83 | 0.77 | 0.90 | 0.84 | 0.70 | 0.82 |
| Osoba and associates [3] | Heterogeneous cancer (535) | 0.68 | 0.54 | 0.56 | 0.73 | 0.68 | 0.86 | 0.80 | 0.82 | 0.65 |
| Ringdal and Ringdal [4] | Heterogeneous cancer (177) | 0.75 | 0.55 | 0.65 | 0.85 | 0.72 | 0.85 | 0.83 | 0.86 | 0.84 |

PF, physical functioning; RF, role functioning; CF, cognitive functioning; EF, emotional functioning; SF, social functioning; QL, a global QOL; FA, fatigue; PA, pain; NV, nausea/vomiting.

*Table 5. Correlation among the EORTC QLQ-C30 scales*

| | | Pearson's correlation coefficient | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | PF | RF | CF | EF | SF | FA | PA | NV | QL |
| Japanese EORTC QLQ-C30 (*n* = 93) | | | | | | | | | | |
| Physical | (PF) | | 0.83 | 0.60 | 0.53 | 0.44 | –0.61 | –0.46 | –0.37 | 0.52 |
| Role | (RF) | | | 0.59 | 0.59 | 0.40 | –0.64 | –0.44 | –0.28 | 0.56 |
| Cognitive | (CF) | | | | 0.78 | 0.61 | –0.70 | –0.56 | –0.45 | 0.58 |
| Emotional | (EF) | | | | | 0.52 | –0.70 | –0.54 | –0.43 | 0.69 |
| Social | (SF) | | | | | | –0.49 | –0.49 | –0.34 | 0.37 |
| Fatigue | (FA) | | | | | | | 0.69 | 0.47 | –0.73 |
| Pain | (PA) | | | | | | | | 0.48 | –0.55 |
| Nausea/vomiting | (NV) | | | | | | | | | –0.39 |
| Global QOL | (QL) | | | | | | | | | |
| Data reported by Aaronson and associates [1] (*n* = 305) | | | | | | | | | | |
| Physical | (PF) | | 0.57 | 0.33 | 0.27 | 0.35 | –0.61 | –0.32 | –0.14 | 0.55 |
| Role | (RF) | | | 0.23 | 0.25 | 0.31 | –0.54 | –0.29 | –0.14 | 0.41 |
| Cognitive | (CF) | | | | 0.44 | 0.35 | –0.38 | –0.34 | –0.12 | 0.32 |
| Emotional | (EF) | | | | | 0.37 | –0.41 | –0.31 | –0.18 | 0.46 |
| Social | (SF) | | | | | | –0.41 | –0.42 | –0.21 | 0.49 |
| Fatigue | (FA) | | | | | | | 0.46 | 0.29 | –0.62 |
| Pain | (PA) | | | | | | | | 0.23 | –0.47 |
| Nausea/vomiting | (NV) | | | | | | | | | –0.26 |
| Global QOL | (QL) | | | | | | | | | |

Aaronson and associates [1]. Compared with patients with lung cancer in the report by Aaronson and associates [1], our patients with lung cancer had a higher stage (stage IV patients/total patients: 29.7% versus 64.8%, respectively) and poorer PS (patients with ECOG PS 3–4/total patients: 6.4% versus 31.4%, respectively). This might have influenced the poorer scores, indicating that it was impossible to show accurately metric equivalence, i.e., norming, from this study [5, 12]. However, the response frequency distributions of both results were alike (Table 2).

In the reliability analysis, Cronbach's alpha coefficient to judge internal consistency was somewhat low in cognitive functioning. Similar results were found in the results of Western studies in patients with advanced lung cancer [1], head and neck cancer [2] and heterogeneous cancer [3, 4] (Table 4).

Regarding item-discrimination, the multitrait scaling analysis confirmed that the Japanese EORTC QLQ-C30 had weak points in the areas of role functioning and cognitive functioning (Table 3). Aaronson and associates and Kaasa and associates reported the same result regarding role functioning [1, 9] (Table 3). From the result of interscale correlations (Table 5), cognitive function correlated with many scales/items of physical functioning, role functioning, emotional functioning, social functioning, global quality of life, fatigue, pain, nausea/vomiting, dyspnoea, sleep disturbance, appetite loss and constipation at more than 0.4 Pearson's correlations, indicating that cognitive function might have a relatively broader conception in Japan. Furthermore, cognitive function was highly correlated with emotional function ($r = 0.78$). These were considered to be the reasons for poor item-discrimination.

*Table 6. Known-group comparison using ECOG performance score (PS)*

| | Patients with PS 0–1 ($n = 47$) | | Patients with PS 2–4 ($n = 46$) | | *t*-test *P* value |
|---|---|---|---|---|---|
| | Mean score | S.D. | Mean score | S.D. | |
| Functioning scales | | | | | |
| Physical | 68.9 | 30.3 | 27.6 | 28.3 | < 0.001 |
| Role | 76.6 | 34.3 | 25.0 | 35.4 | < 0.001 |
| Cognitive | 77.3 | 20.4 | 50.6 | 28.0 | < 0.001 |
| Emotional | 83.9 | 18.8 | 58.6 | 28.5 | < 0.001 |
| Social | 74.1 | 21.9 | 56.9 | 33.1 | 0.006 |
| Global quality of life | 54.6 | 20.1 | 35.9 | 26.1 | < 0.001 |
| Symptom scales and/or items | | | | | |
| Fatigue | 35.3 | 23.3 | 61.6 | 30.3 | < 0.001 |
| Pain | 27.2 | 27.2 | 45.8 | 35.9 | 0.010 |
| Nausea and vomiting | 10.6 | 18.9 | 24.4 | 29.1 | 0.012 |
| Dyspnoea | 20.6 | 22.6 | 40.2 | 37.4 | 0.004 |
| Sleep disturbance | 21.3 | 26.4 | 41.2 | 32.9 | 0.003 |
| Appetite loss | 19.9 | 27.5 | 49.0 | 36.0 | < 0.001 |
| Constipation | 18.4 | 26.8 | 37.3 | 39.2 | 0.012 |
| Diarrhoea | 5.2 | 13.9 | 16.0 | 27.4 | 0.023 |
| Financial impact | 18.1 | 21.6 | 27.0 | 34.3 | 0.157 |

*Table 7. Longitudinal clinical validity with Karnofsky Performance Status Scale (KPS)*

| | Improved KPS ($n = 22$) | | Deteriorated KPS ($n = 24$) | | |
|---|---|---|---|---|---|
| Mean score (S.D.) | First estimation | Later estimation | First estimation | Later estimation | *P* value* |
| Functioning scales | | | | | |
| Physical | 47.3 (32.4) | 47.3 (36.3) | 49.2 (32.8) | 41.7 (36.3) | 0.614 |
| Role | 52.3 (39.3) | 56.8 (44.4) | 52.1 (42.9) | 37.5 (42.4) | 0.247 |
| Cognitive | 68.9 (27.3) | 65.9 (26.9) | 66.7 (23.5) | 51.4 (31.1) | 0.011 |
| Emotional | 77.6 (19.1) | 76.2 (26.3) | 76.5 (20.6) | 65.8 (26.3) | 0.147 |
| Social | 72.8 (27.5) | 66.7 (30.9) | 68.7 (25.7) | 58.3 (36.1) | 0.080 |
| Global quality of life | 54.4 (22.5) | 57.6 (18.0) | 44.4 (23.0) | 39.2 (26.2) | 0.544 |
| Symptom scales and/or items | | | | | |
| Fatigue | 41.1 (20.2) | 35.4 (21.3) | 49.1 (25.0) | 60.3 (26.6) | 0.030 |
| Pain | 31.8 (24.6) | 22.5 (21.7) | 31.7 (31.7) | 37.5 (29.2) | 0.144 |
| Nausea and vomiting | 18.2 (22.4) | 10.6 (17.5) | 16.0 (23.3) | 27.8 (32.5) | 0.034 |
| Dyspnoea | 25.8 (27.1) | 18.2 (19.9) | 29.2 (26.6) | 37.5 (33.1) | 0.139 |
| Sleep disturbance | 27.3 (24.4) | 25.5 (14.2) | 26.4 (27.8) | 31.9 (30.3) | 0.663 |
| Appetite loss | 28.8 (29.6) | 31.8 (31.7) | 36.1 (31.0) | 55.6 (32.1) | 0.189 |
| Constipation | 25.8 (37.0) | 24.2 (27.6) | 29.2 (28.3) | 34.7 (33.3) | 0.624 |
| Diarrhoea | 9.1 (18.4) | 10.6 (18.9) | 10.2 (18.2) | 22.2 (28.9) | 0.166 |
| Financial impact | 23.9 (31.1) | 33.0 (38.5) | 19.1 (25.7) | 26.4 (34.0) | 0.440 |

*Based on repeated-measures ANOVA.

22 patients had an improved Karnofsky performance status scale (KPS) at more than one point during the investigation, KPS deteriorated in 24 patients at more than one point and KPS was unchanged in 22 patients (data are not shown). The remaining patients were investigated once ($n = 15$) and due to missing KPS ($n = 10$).

ECOG PS and KPS were judged by a third person such as a physician, and these parameters were thought to be internationally validated. For estimation of clinical validity of the Japanese QLQ-C30, a known-group comparison using ECOG PS was satisfied (Table 6). KPS was used in the preliminary investigation of responsiveness to change in health status, and longitudinally clinical validity with KPS was found for the scales of cognitive functioning, fatigue and nausea/vomiting (Table 7). Aaronson and associates reported that the longitudinally clinical validity with ECOG PS was found for the scales of physical functioning, role functioning, global QOL scale, fatigue and nausea/vomiting with European people. Our patients had poorer PS and were all hospitalised. This might reflect the difference in the scales of both physical and role functioning.

In correlations between global QOL and the other scales (Table 5), there were differences between Japanese and European results. Cognitive functioning moderately correlated with global QOL, as mentioned above, and social functioning only poorly correlated with global QOL for Japanese. The different structure of QOL may be one of the reasons for the cultural division between Japan and the West. A multiple regression equation, when the global QOL scale is the criterion variable and the other scales are explanatory variables was used to assess the QOL structure for Japanese patients with lung cancer, and it showed that emotional functioning (regression coefficient = 0.38) and fatigue (regression coefficient = 0.42) were significant explanatory variables for the global QOL scale, but, to our knowledge, there have been no published reports in the West on this issue. The result of the multiple regression equation indicates that international comparisons of the scales on the QLQ-C30 are possible, but that, as the QLQ-C30 is not designed to be aggregated by simple summation of all the item-scores, comparison of the total score is not valid.

In summary, the Japanese QLQ-C30 has a weak scale of role functioning in terms of item discriminative validity. It also has a weak scale of cognitive functioning in items of discriminative validity and internal consistency. However, the other scales constituting QOL in the English-speaking culture were extracted and found to be valid in Japan, indicating its possible usefulness as an instrument that is universally applicable across cultures. Now the EORTC QLQ-C30 can be used in international phase III studies, for example, among Europe, North America and Japan. In those patients who will be selected by the strict entry criteria in the phase III study, the metric equivalence will be accurately evaluated [5, 12].

1. 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**, 365–375.
2. Bjordal K, Kaasa S. Psychometric validity of EORTC core quality of life questionnaire. 30 item version, a diagnosis specific module for head and neck cancer patients. *Acta Oncol* 1992, **31**, 311–321.
3. Osoba D, Zee B, Pater J, Warr D, Kaizer L, Latreille J. Psychometric properties and responsiveness of the EORTC quality of life questionnaire (QLQ-C30) in patients with breast, ovarian and lung cancer. *Qual Life Res* 1994, **3**, 353–364.
4. Ringdal GI, Ringdal K. Testing the EORTC quality of life questionnaire on cancer patients with heterogeneous diagnoses. *Qual Life Res* 1993, **2**, 129–140.
5. Bullinger M, Power MJ, Aaronson NK, Cella DF, Anderson RT. Creating and evaluating cross-cultural instruments. In Spilker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials*, 2nd edn. Philadelphia, Lippincott-Raven, 1996, 659–668.
6. Hui C, Triandis HC. Measurement in cross-cultural psychology: a review and comparison of strategies. *Cross Cultural Psychol* 1985, **16**, 131–152.
7. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In MacLeod CM, ed. *Evaluation of Chemotherapeutic Agents*. New York, Columbia University Press, 1949, 191–205.
8. Sprangers MAG, Cull A, Bjordal K, Groenvold M, Aaronson NK. The European Organization for Research and Treatment of Cancer approach to quality of life assessment: Guidelines for developing questionnaire modules. *Qual Life Res* 1993, **2**, 287–295.
9. Kaasa S, Bjordal K, Aaronson N, *et al.* The EORTC Core Quality of Life Questionnaire (QLQ-C30): Validity and reliability when analyzed with patients treated with palliative radiotherapy. *Eur J Cancer* 1995, **31A**, 2260–2263.
10. Hays RD, Hayashi T, Carson S. *A User's Guide for the Multitrait Analysis Program*. Santa Monica, California, Rand Corporation, 1988.
11. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951, **16**, 297.
12. Anderson RT, Aaronson NK, Leplege AP, Wilkin D. International use and application of generic health-related quality of life instruments. In Spilker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials*, 2nd edn. Philadelphia, Lippincott-Raven, 1996, 613–632.

**Acknowledgements**—Special thanks to Mr M. Takahashi and Mr Y. Matsushita for assistance. This work was supported by grants from the Ministry of Health and Welfare of Japan.