Pergamon

European Journal of Cancer, Vol. 34, No. 6, pp. 767–769, 1998
© 1998 Elsevier Science Ltd. All rights reserved
Printed in Great Britain
0959-8049/98 $19.00+0.00

PII: S0959-8049(98)00073-2

Editorial

Assessing the Quality of Life of Patients with Cancer: East Meets West

N.K. Aaronson

[1]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

CLINICAL RESEARCH in oncology is increasingly characterised by multicentre efforts on an international scale. This has the advantage of improving the efficiency with which clinical trials can be conducted via rapid accrual of patients and also facilitates the registration of new, effective therapies across national boundaries and health care systems. In order to carry out such international studies it is essential that the outcomes of interest are standardised across participating countries. This presents few, if any, problems when focusing on such classical outcomes as tumour response and survival. However, when employing treatment outcomes of a more subjective nature, including patients' self-reported symptom experience, perceived health status and quality of life (QL), cross-cultural issues become of paramount importance. It is essential that the questions that we ask and the responses that we elicit from our patients maintain the same meaning when translated into different languages and applied in different cultural settings. Without such cross-cultural equivalence, the interpretation of patient-based data in multinational clinical trials is at best problematic and at worst impossible.

The past 15 years have witnessed major advances in the development of practical, reliable and valid questionnaires for assessing the perceived health status and QL of patients with cancer. These include the Rotterdam Symptom Checklist (RSCL), the Functional Living Index—Cancer (FLIC), the Cancer Rehabilitation Evaluation System (CARES), the Functional Assessment of Cancer Therapy (FACT) and the European Organization for Research and Treatment of Cancer core QL questionnaire (the QLQ-C30). These questionnaires have proven their usefulness in documenting the symptom and functional health experience of patients with a range of cancer diagnoses and, of perhaps greater importance, they have been applied successfully within the demanding setting of phase II and phase III clinical trials. The FLIC, the CARES and the FACT were originally developed for use among English-speaking patients in North America. The RSCL, although originally developed in the Netherlands, quickly enjoyed widespread use in the U.K. After successful introduction in their countries of origin, these questionnaires have been translated into a number of additional languages and reports on their use in non-English-speaking research settings are beginning to appear in the literature.

In contrast, the EORTC QLQ-C30 was, from the outset, developed with an eye toward cross-cultural application. This is understandable given the multinational composition of the EORTC itself, with representation from the large majority of Western European countries, as well as an increasing number of countries from Eastern Europe and even farther afield. Rather than taking a *sequential* approach to questionnaire construction (i.e. development and testing in one country, followed by translation and application in additional countries), the EORTC employed a *parallel* approach, with input from health professionals and patients drawn from a range of countries and cultures at each step in the questionnaire development process (i.e. identifying relevant QL domains, item construction, translation and pretesting and large scale field-testing of psychometric properties).

Version 1.0 of the QLQ-C30 was originally tested among lung cancer patients accrued from centres in 13 countries [1]. The study results supported the reliability and validity of the questionnaire, both for the sample as a whole and for subgroups of patients from Northern Europe, Southern Europe and from English-speaking countries (i.e. Australia, Canada, and the U.K.). Subsequent studies have confirmed the psychometrics of the questionnaire when employed with a wider range of diagnostic groups [2]. More recently, the QLQ-C30 has undergone a number of minor revisions which have resulted in increased reliability and precision of several of its scales [3].

The QLQ-C30 is currently available in 25 languages, including not only the major languages of Western and Eastern Europe, but also Chinese, Hebrew and Japanese (with translations pending in Hindi, Korean, Malaysian and Sotho). These various language adaptations of the questionnaire have been generated with the use of a standardised, iterative, forward-backward translation protocol and have undergone centralised quality control. To date, however, the psychometric evaluation of the QLQ-C30 has been limited to English-speaking and European cultures and languages.

For this reason, the report by Kobayashi and colleagues in this issue of the *European Journal of Cancer* (pp. 810–815) on

---

Received and accepted 19 Feb. 1998.

the psychometric performance of the questionnaire when employed among cancer patients in Japan is of particular interest and importance. The study focused on patients with lung cancer, as was the case in the original international field study of the QLQ-C30. However, in contrast to the international study, the Japanese patients were in a more advanced stage of disease, were hospitalised and had a poorer initial performance status.

Despite the fact that the patients in this study were rather ill, the overall compliance rate was satisfactory. Approximately 90% of eligible patients completed the initial QLQ-C30 at or around the time of hospital admission and 83% completed all planned questionnaires during their hospital stay (i.e. every 2 weeks). This emphasises the fact that, given sufficient organisational and staff support, high rates of compliance can be achieved, even in QL investigations of patients with advanced disease. At the same time, this study points to the practical limits of such QL investigations. Whereas between 81 and 99% of the planned questionnaires were completed by patients with an ECOG performance status score of 0–3, this figure fell dramatically to 13% for patients with an ECOG score of 4. This problem of censored (non-random missing) QL data is of great concern in that it can seriously bias study findings and can lead to erroneous conclusions. A number of approaches have been proposed for dealing with this problem, ranging from the use of proxy respondents (e.g. health care providers or family members) in studies where high rates of attrition are anticipated, to the use of *post-hoc* imputation methods for estimating missing QL scores for patients lost to follow-up. To date, however, no consensus has been achieved as to how best to deal with this all too common problem.

Kobayashi's group employed a standard set of psychometric analyses to examine the reliability and validity of the QLQ-C30. Multitrait scaling analysis provided strong support for the scale structure of the questionnaire, with scaling errors observed in only 2.1% of the tests carried out (as compared with 4% in the international field study). Internal consistency reliability was above the 0.70 criterion recommended for group comparisons for 7 of the 8 multi-item scales, the exception being the cognitive functioning scale. In fact, the reliability coefficients obtained in the Japanese study were consistently higher than those reported in previous studies.

While the scale structure of the QLQ-C30 was confirmed in this study, it is noteworthy that the correlations observed between the scales were, on average, substantially higher than those found in the international field study (mean correlation across all scales of 0.54 versus 0.36, respectively). Of particular importance is the relatively high correlation observed between scales assessing physical versus psychosocial health domains (e.g. $r=0.60$ and 0.53 between the physical functioning scale and the cognitive and emotional functioning scales, respectively). In part, this might be explained by the fact that the QLQ-C30 scores for the Japanese sample tended to show more variability than those of the European sample. All else being equal, increased score variance will result in higher interscale correlations. At the same time, one cannot rule out a more substantive, cross-cultural explanation. It may be, for example, that the Japanese culture adheres less strongly to a Cartesian view of health than is the case in the West; that the illness experience of Japanese patients is of a more holistic nature. One would then expect to observe a higher degree of covariation between the physical and psychosocial dimensions of health. From a measurement perspective, however, the implication of these findings should not be overstated. An interscale correlation of 0.60 indicates that only approximately one-third of the score variance is shared between scales. In other words, these findings do not challenge the distinctiveness of the health domains assessed by the QLQ-C30.

A series of so-called 'known group' comparisons yielded consistent support for the validity of the QLQ-C30. Statistically significant differences between patients with low versus high ECOG performance status were observed for all the scales. Although not reported in the paper, the magnitude of these differences, as expressed in terms of standardised scores (i.e. effect size estimates) was substantial. The average effect size across the 5 functional scales and the overall quality of life was 1.0 (range from 0.63–1.5), indicating differences between groups in the order of more than 0.5 to 1.5 standard deviations (S.D.). For the symptom scales and items, effect sizes ranged from 0.5 to greater than 1.0. As a rule of thumb, effect sizes greater than 0.5 are considered to be moderate, and 0.8 or greater to be large.

The analysis of QLQ-C30 change scores, comparing patients whose Karnofsky Performance Status (KPS) had improved with those whose KPS had deteriorated, yielded only a few statistically significant results (for cognitive functioning, fatigue, and nausea and vomiting). In part, this may have been due to the relatively small sample size available for the repeated measures analysis, as well as to the fact that a shift of only 1 scale level on the KPS was used as the criterion for change in performance status. Nevertheless, irrespective of statistical significance levels, the pattern of results was quite consistent. Patients whose KPS had improved tended to maintain relatively stable levels of functioning and to experience some symptom relief. In contrast, those patients whose KPS had declined, reported a parallel worsening of functioning and an increase in symptom levels.

Finally, the authors report the results of a multiple regression analysis in which the global QL scale was regressed on the other scales and single item measures of the QLQ-C30. Only two scales, emotional functioning and fatigue, were found to be associated significantly with overall QL. This is striking given that, in version 1.0 of the QLQ-C30, the two-item global QL scale places relatively greater emphasis on physical health (i.e. one item asks patients to rate their overall physical condition; the other to rate their overall quality of life). In a similar analysis of the same version of the questionnaire based on data from a heterogeneous sample of Norwegian cancer patients, a larger number of scales was found to be associated significantly with global QL at the multivariate level, including physical and role functioning, emotional functioning, cognitive functioning and fatigue. Together they accounted for approximately 50% of the variance in the global QL scores [4]. It is likely that the QL domains that contribute most strongly to explaining overall QL scores will vary as a function of diagnosis, stage of disease, treatment status, as well as sociocultural and sociodemographic variables. Clearly though, both physical and psychosocial variables factor into the equation.

In summary, the findings of Kobayashi and colleagues suggest that the EORTC QLQ-C30 is a robust instrument which can be employed successfully in assessing the QL of patients in non-Western cultures. The problems observed with certain scales have been noted previously and have led to revisions of the questionnaire. For example, version 2.0 incorporated changes in the role functioning scale that dramatically improved its reliability [3] and the recently released version 3.0 introduces a slightly modified physical functioning scale that increases substantially the number of response levels available. Such fine-tuning of extant QL instruments is both necessary and desirable in order to increase the precision of measurement.

It is important to emphasise that Japan is a highly industrialised, economically successful country. Additional research is needed to determine the feasibility and appropriateness of applying measures such as the QLQ-C30 in Third World countries, as well as among economically and socially disadvantaged minority groups in Western societies. Reports of recent successes in adapting the FACT [5] and the SF-36 Health Survey [6] for use among minority populations in the United States and of the World Health Organization in developing a QL instrument (the WHOQOL) for use in countries as culturally diverse as India, Thailand and Zimbabwe [7] bode well for the standardisation of QL and health status assessment for use in clinical research on a global scale.

1. Aaronson NK, Ahmedzia 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.
2. Aaronson NK, Cull AM, Kaasa S, Sprangers MAG. The European Organization for Research and Treatment of Cancer (EORTC) modular approach to quality of life assessment in oncology: an update. In Spilker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials*, 2nd edition. Philadelphia, Lippincott-Raven, 1996, 179–190.
3. Osoba D, Aaronson N, Zee B, Sprangers M, te Velde A. Modification of the EORTC QLQ-C30 (version 2.0) based on content validity and reliability testing in large samples of patients with cancer. *Quality Life Res* 1997, **6**, 103–108.
4. Ringdal GI, Ringdal K. Testing the EORTC quality of life questionnaire on cancer patients with heterogenous diagnoses. *Quality Life Res* 1993, **2**, 129–140.
5. Cella DF, Bonomi AE. The Functional Assessment of Cancer Therapy (FACT) and Functional Assessment of HIV Infection (FAHI) quality of life measurement system. In Spilker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials*, 2nd edition. Philadelphia, Lippincott-Raven, 1996, 203–225.
6. Johnson PA, Goldman L, Orav EJ, Garcia T, Pearson SD, Lee TH. Comparison of the Medical Outcomes Study Short-Form 36-Item Health Survey in black patients and white patients with acute chest pain. *Med Care* 1995, **33**, 145–160.
7. The WHOQOL Group. The World Health Organisation quality of life assessment (WHOQOL): position paper from the World Health Organisation. *Soc Sci Med* 1995, **41**, 1403–1409.