PERGAMON

European Journal of Cancer 37 (2001) 1331–1334

European Journal of Cancer

www.ejconline.com

Editorial Comment

Interpreting quality of life data: population-based reference data for the EORTC QLQ-C30

P.M. Fayers [a,b,*]

[a] *Department of Public Health, Aberdeen University, Scotland, UK*
[b] *Unit of Applied Clinical Research, Norwegian University of Science and Technology (NTNU), Norway*

Received 1 March 2001; accepted 13 March 2001

Although the European Organization for Research and Treatment of Cancer (EORTC) core 30-item Quality of Life Questionnaire (QLQ-C30) is widely used in cancer patients, as with all health-related quality of life (QoL) instruments there are issues of interpretation. What does it mean if, on a scale from 0 to 100, a cancer patient reports a score of 50 for overall QoL? Is an improvement to 55 worthwhile? Various approaches have been used to aid interpretation [1], and one method is to produce tables of 'normative' or 'population-based reference values', based upon the general population. Normative data ('norms') make it possible to compare the score of a patient against the average score within the population. Norms, being usually based upon a large random sample from a national population, also allow the impact of covariates such as age and gender to be explored. Since QoL scores vary markedly with both these factors, it will frequently be important to make allowance for them when interpreting the results from clinical trials or other studies, or when examining the scores for individual patients.

It is, therefore, interesting to read the paper by Schwarz and Hinz [2], providing reference data for the EORTC QLQ-C30 in a random sample of the German population. This complements the existing data for the three Scandinavian countries — Klee and colleagues (1997) in Danish women [3], Hjermstad *et al.* (1998) in Norway [4] and Michelson *et al.* (2000) in Sweden [5]. Of particular interest is the comparison of these four surveys. Since they were all conducted in Northern countries, one might perhaps anticipate that differences should be minor — certainly, one might be less surprised to find substantial variation between Germany and the Mediterranean countries, for example, but there are arguably lesser differences culturally and linguistically between the countries of Germany and Scandinavia.

Thus it is surprising to note the extent of the differences in QoL between Germany and Scandinavia. Some of these differences are discussed by Schwartz and Hinz. The average 'global QoL' scores for males are shown in Fig. 1(a), with females in Fig. 1(b) (Danish data was only for females, and must be interpreted with caution because an earlier version of the QLQ-C30 with different wording was used). For both sexes, the German results show a consistently steeper decline with age, ending at far lower scores in older patients. Another example of the differences is the fatigue scale (Fig. 2a and b).

Why are there such large discrepancies? One possibility is that there are translation errors or inadequacies in the German version, either in terms of the response categories or the overall question itself. This seems unlikely for three reasons. Firstly, the translation procedure that is used for the QLQ-C30 is rigorously defined, and involves iterated forward and backward translations by independent translators who are each native speakers of their target language, followed by field-testing and debriefing of patients [6]. Secondly, as Schwartz and Hinz note, there are substantial differences between Germany and Scandinavia for many of the scales, especially the symptom items and scales in which Germans report fewer problems. It seems unlikely that translation errors would arise for so many items. Thirdly, and perhaps most convincingly of all, similar patterns for overall QoL can be seen with other instruments. Reference data is available for the SF-36 [7], and this shows similar patterns. For general health, Norway, Denmark and Sweden score 76.2, 75.0 and 75.0, respectively, while Germany was appreciably lower at 64.9. This is an even larger difference than Schwartz and Hinz found for the global QoL scale of the QLQ-C30.

* Tel.: +44-1224-559-573; fax: +44-1224-662-994.
E-mail address: p.fayers@abdn.ac.uk (P.M. Fayers).

PII: S0959-8049(01)00127-7

Another possible explanation for differences between the countries is selection bias. Although each study purported to be a random sample of the general population, perhaps one or another was more selective than the rest. This, too, seems unlikely given the similar pattern shown by the independent surveys of the SF-36. In addition, the patient participation rates in the four QLQ-C30 surveys were broadly similar, with between 68 and 78% of patients responding to the questionnaires.

A third possibility is that the observed scores reflect actual differences in the health of the countries. The general population is certainly not the same as a 'healthy population' — the Norwegian study found that 68% of the subjects reported chronic health problems, and similarly in the Swedish study 20% had hypertension, 18% arthritis, 18% reduced limb function, 12% allergies, 8% angina and so on. Doubtless diet, obesity and health style, as well as the prevalence of serious chronic conditions will vary across countries. Thus it seems very likely that there could be national differences in overall health and in QoL. One might speculate that national differences in overall health and physical fitness would become more prominent with age, as appears to be the case in Fig. 1.

Cultural effects offer a fourth explanation. Thus it might be possible for different communities to perceive or describe their QoL differently, even although as far as an outsider can tell there is no discernable difference in their well-being. This could arise if their expectations were different, for example if persons in one country had higher expectations for being able to pursue sporting activities as a major part of their daily life. In surveys such as these, cultural effects are difficult to separate from the other causes such as differing general health, but could well explain the patterns of Fig. 2. Because this is an important issue, the EORTC Quality of Life Group (QLG) has initiated a major project to explore cultural differences; the QLG, through the EORTC randomised trials and the work of the QLG members, has assembled a database of over 10 000 patients from European countries for an analysis of cultural effects.

Perhaps the reason the German data seems so different is a combination of the third and fourth explanations; possibly the disparity is explained partly by real differences and partly by cultural effects. It would be interesting to explore comparable data from other countries, and investigators should be encouraged also to present details of chronic diseases in addition to other baseline data.

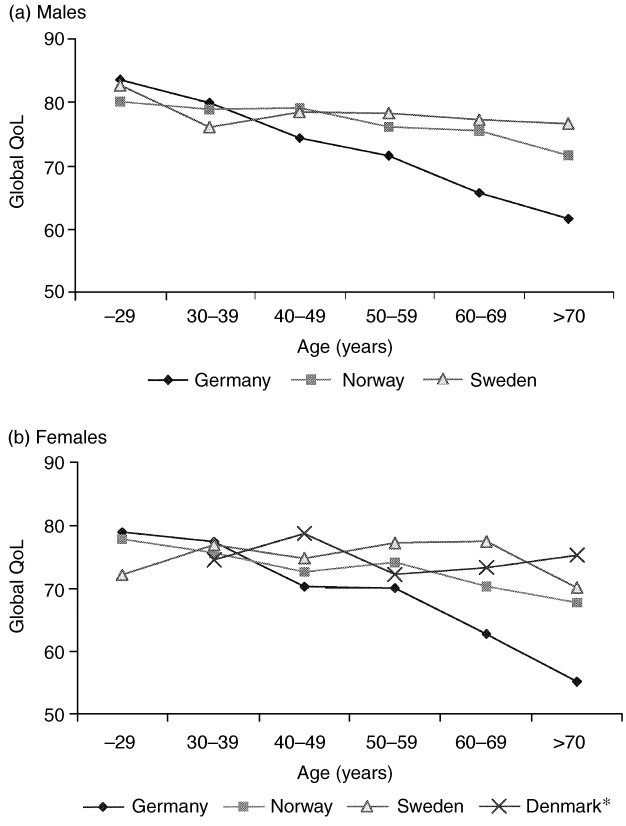

Fig. 1. (a) Mean European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 global Quality of Life (QoL) score, males; (b) mean EORTC QLQ-C30 global QoL score, females (*Danish data used EORTC QLQ-C30 version 1.0).

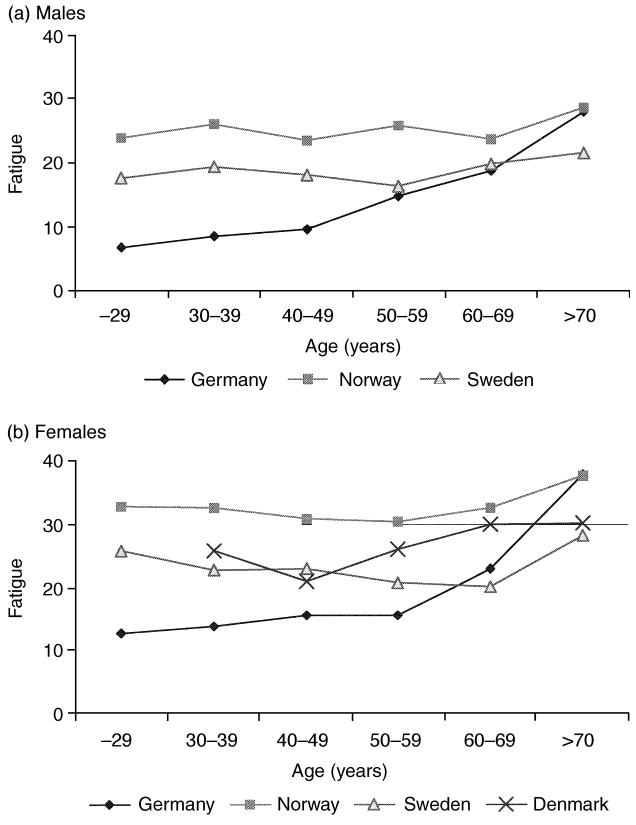

Fig. 2. (a) Mean EORTC QLQ-C30 fatigue score, males; (b) mean EORTC QLQ-C30 fatigue score, females.

How should one use reference values? The first point to note is that the general population has a far from perfect QoL. Perhaps a modest initial target in patient management might be to try to ensure that patients' QoL is no worse than the average of the age and gender-matched population. Thus reference values may serve as a target, and patient scores can be interpreted in terms of either their absolute difference from the reference value or the percentage reduction. However, with seriously ill patients, the population averages might represent a clearly unattainable target. A more meaningful approach could be to translate patients' scores into their percentile position. For example, from the Norwegian reference data we find that just under 10% of females aged 50–59 years report global QoL scores of 33 or less. Then, when an individual female patient of this age scores 25, we could say 'This patient's QoL is as poor as that of the lowest 10% of the population'. Such statements may be more easily interpretable than citing scores and normative mean values. In addition, as Rose and colleagues observe, the use of percentiles overcomes the problems of QoL scores frequently following a highly non-normal asymmetric distribution [8]. Thus the lower 10th percentiles of the population data would provide useful reference points. However, percentiles are of limited applicability with four-point single-items and those scales that have few categories. Instead, the percentages of patients and the general population in the poorest categories can be contrasted with each other. For example, 11% of males and 22% of females in the Norwegian reference data reported 'quite a bit' of insomnia, and one can compare this with the percentage of patients experiencing similar levels of problems. Unfortunately, percentiles are not available in any of the papers discussed, which present mean scale scores and standard deviations [2–5].

In these ways, norms can be used as reference values against which a patient's score may be compared, or against which a group mean score or percentage can be contrasted. Norms can also be used as target values when managing patients. But they do not answer such questions as 'How important to a patient is a change of (say) 10 in their QoL score?', or in a clinical trial, 'If the mean scores of the two treatment groups differ by 10, is this clinically important and worthwhile?'

To attempt to answer these questions, a number of other approaches have been devised [1]. Statistical significance is definitely *not* an appropriate method for defining clinical significance, because for any given change or treatment effect the P values are associated with the sample size. One method that is appropriate is the concept of 'minimal clinically important difference' or the 'subjective significant change'. The objective is to establish the smallest difference that patients notice and perceive as moderately beneficial. Therefore, if no troublesome side-effects occur, and if there are no excessive cost implications, a potential improvement of this order should influence the choice of patient management. Various investigators, using a variety of QoL instruments, have found that an absolute change of 10% or more in QoL scores corresponds on average to patients' description of 'a moderate change', and 5–10% represents a small change, and Osoba and colleagues have shown that this applies to the QLQ-C30 [9]. This conclusion is also consistent with the results from other approaches, such as the use of Cohen's 'effect size' statistics [1].

The interpretation of QoL results remains essentially qualitative. Clinical significance is subjective, and is a matter of opinion. The values and opinions of individual patients will differ, as will the opinions of the treating clinician and those of society in general. Thus, for a QoL measurement scale, it is unlikely that a single threshold value will be universally accepted as a cut-off point that separates clinically important changes from trivial and unimportant ones. It is also likely that patients may consider changes in some aspects of QoL to be more important than others, and a change of, say, five points on one scale may be as clinically important as a change of 20 on another; a change from 'a little' to 'very much' may be far more serious for some symptoms than others.

In conclusion, therefore, population-based reference values provide an important aid to the interpretation of QoL scales, and the paper of Schwartz and Hinz makes a useful contribution to this knowledge. The differences between the German and Scandinavian studies indicate a need for further national reference data sets from other countries. It is also recommended that percentiles be reported, in addition to mean values.

References

1. Fayers P, Machin D. *Quality of Life: Assessment, Analysis and Interpretation*. Chichester, J. Wiley & Sons, 2000.
2. Schwarz R, Hinz A. Reference data for the quality of life questionnaire EORTC QLQ-C30 in the general German population. *Eur J Cancer* 2001, **37**, 1345–1351.
3. Klee M, Groenvold M, Machin D. Quality of life of Danish women: population-based norms for the EORTC QLQ-C30. *Quality of Life Research* 1997, **6**, 27–34.
4. Hjermstad MJ, Fayers P, Bjordal K, Kaasa S. Health-related quality of life in the general Norwegian population assessed by the European Organization for Research and Treatment of Cancer quality-of-life questionnaire: the QLQ-C30(+3). *J Clin Oncol* 1998, **16**, 1188–1196.
5. Michelson H, Bolund C, Nilsson B, Brandberg Y. Health related quality of life measured by the EORTC QLQ-C30 — reference values from a large sample of the Swedish population. *Acta Oncologica* 2000, **39**, 477–484.
6. Cull A, Sprangers M, Bjordal K, Aaronson N, on behalf of the EORTC Quality of Life Study Group. *EORTC Quality of Life Study Group Translation Procedure*. EORTC, Brussels, 1998.

7. Gandek B, Ware JE, Aaronson NK, *et al*. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA project. *J Clin Epidemiol* 1998, **51**, 1149–1158.
8. Rose MS, Koshman ML, Spreng S, Sheldon R. Statistical issues in the comparison of health-related quality of life in diseased patients to published general population norms: problems and solutions. *J Clin Epidemiol* 1999, **52a**, 405–412.
9. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998, **16**, 139–144.