Pergamon

European Journal of Cancer, Vol. 34, No. 9, pp. 1381–1389, 1998

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

PII: S0959-8049(98)00136-1

Original Paper

Using Reference Data on Quality of Life—the Importance of Adjusting for Age and Gender, Exemplified by the EORTC QLQ-C30 (+3)

M.J. Hjermstad,[1] P.M. Fayers,[2] K. Bjordal[3] and S. Kaasa[2,4]

[1]The Norwegian Cancer Society and Department of Medical Oncology and Radiotherapy, The Norwegian Radium Hospital, Montebello, N-0310,Oslo; [2]Unit for Applied Clinical Research, Faculty of Medicine, University of Science and Technology, Trondheim; [3]Department of Oncology, Trondheim University Hospital; and [4]Palliative Medicine Unit, Department of Oncology, Trondheim University Hospital, Trondheim, Norway

Interpretation of health related quality of life (HRQOL) results in cancer patients is facilitated by knowledge of the levels of HRQOL in the general population. However, direct comparisons can be misleading unless age and gender are considered. We demonstrate the derivation of age- and gender-specific 'expected' values from population reference values by means of simple calculations. This survey included 3000 randomly selected Norwegians above 18 years of age who received the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ-C30 (+3)) by mail. 1965 responses from 2,892 eligible persons (68%) were received. The population was divided into six disease groups based on self-reported health problems. The observed mean scale scores of the different groups deviated greatly from those obtained in the general population. The score for physical function, for example, was 72 for cancer patients and ranged from 73.3 to 82.5 in other disease groups, as opposed to 89.9 in the general population and 98.9 in those with no health problems. The range for one of the quality of life (QOL) scales was 57.7 to 84.7 compared with 73.7 in the general population. Expected mean scores for the patient groups were computed from the reference values, based on the concept of equivalence of age and gender. The differences between the observed mean scores and the reference values were strongly mediated by this method. The expected scores for physical function then ranged from 83.3 to 93.1 and from 70.3 to 75 for the QOL scale. The impact of age and gender on the reference data from the EORTC QLQ-C30 (+3) obtained in a general population shows that these variables must be considered when interpreting data on HRQOL for cancer patients. The demonstration of how to generate mean values which are adjusted according to the age and gender distribution of a population should increase the usefulness of this questionnaire among clinicians. © 1998 Elsevier Science Ltd. All rights reserved.

Key words: health related quality of life, reference data, clinical importance, EORTC QLQ-C30
Eur J Cancer, Vol. 34, No. 9, pp. 1381–1389, 1998

INTRODUCTION

CANCER-SPECIFIC questionnaires, like the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ-C30) [1], the Rotterdam Symptom Checklist (RSCL) [2], the Functional Living Index Cancer (FLIC) [3], the Functional Assessment of Cancer Therapy Scale (FACT) [4] and the Cancer Rehabilitation Evaluation System (CARES) [5], have been developed for assessment of health related quality of life (HRQOL) in various cancer populations and for comparisons across diagnoses and stages of disease.

Reference data from the general population on generic HRQOL questionnaires like the Sickness Impact Profile (SIP) [6] and the SF-36 in particular [7], have been derived

---

Correspondence to M.J. Hjermstad.

Received 17 Oct. 1997; revised 13 Feb. 1998; accepted 25 Mar. 1998.

in multiple studies [6–10]. Hardly any such data are available for cancer-specific questionnaires. The RSCL reports data from healthy controls when measuring HRQOL in different groups of patients [2], while the EORTC QLQ-C30 was employed in a Danish study in order to obtain population values among women [11]. The authors of CARES have developed a database constructed from data obtained in cancer patients free of disease [12]. We have reported reference data for the EORTC QLQ-C30 (+3) in a Norwegian population sample [13]. In this study, men reported better functioning than women, as well as superior scores on the quality of life (QOL) scales, with a gradual decrease in function with increasing age for both genders. Women reported more symptoms than men, also increasing with older age. Females across all ages indicated more problems compared with males. These trends were similar to those observed in studies employing generic questionnaires, like SIP and SF-36 [6–10, 14]. Our results regarding the impact of age and sociodemographic variables on HRQOL, such as education and living situation, were consistent with those in Denmark [11].

We now demonstrate how to interpret scores from the EORTC QLQ-C30 obtained with different patient groups and how to compare these with our previously reported reference data (Appendix 1). Clinicians should take into consideration that impairments not attributable to cancer may be common among cancer patients as well as in the general population of similar age. Age and gender itself also have a profound impact on HRQOL. By means of simple calculations, age and gender adjusted values can be derived. This paper illustrates a general technique in which clinicians may apply our approach to compare our previously published general population data [13] against their own sets of data for diseased patients. This method should be considered an important tool for clinicians when interpreting scores from patients.

PATIENTS AND METHODS

The EORTC QLQ-C30 (+3) questionnaire

The EORTC QLQ-C30 (version 1.0) is a 30 item core questionnaire intended for assessment of HRQOL among cancer patients, developed by the EORTC Quality of Life Study Group. The first version of this questionnaire has been validated in cancer patients and cross culturally tested in more than 24 countries [1, 15–20]. In order to assess symptoms related to particular concerns, it is recommended that the core questionnaire is supplemented by disease-specific modules of which there now exist five official ones for use with colorectal, head and neck, breast, lung and oesophageal cancers [21–25]. We used an interim version of the QLQ-C30 containing three additional test items, the EORTC QLQ-C30 (+3). These three items are now implemented in the next version of the questionnaire, the EORTC QLQ-C30 (version 2.0), replacing the original role function scale and one of two items in the global QOL scale.

Population sample and methods

The sample was obtained by a random draw from the adult Norwegian population according to standard procedures of The Office of the National Register. It consisted of 3000 people and reflected the age and gender distribution of the adult Norwegian population. The EORTC QLQ-C30 (+3), a questionnaire about sociodemographic data and 14 questions about past or present illness and health were sent out by mail. The dichotomous health questions ('yes' or 'no') were identical to those recommended when using the SF-36 [7]. They address prevalent health problems in the population: hypertension, congestive heart failure, myocardial infarction, diabetes, cancer, chronic allergies, arthritis, sciatica or chronic back problems, visual impairment, hearing impairment, chronic obstructive lung problems, chronic skin problems, paralysed/weak limb and other unspecified, chronic medical conditions.

The population was divided into six groups, depending on their health problems. Many subjects had several concurrent health problems. When examining one problem at a time in order to make mutually exclusive groups, it was evident that some of the groups became very small (*n* ranging from 2 to 116 people). Consequently, six groups were constructed based on conceptually similar diseases which are often seen as concomitant problems. Group I ($n=277$) consisted of people suffering from congestive heart failure, myocardial infarction and chronic obstructive lung problems, while group II ($n=694$) included arthritis, sciatica and paralysed/weak limb. Group III ($n=339$) encompassed people with visual or hearing impairments. The chronic disease group, group IV ($n=779$), included those who had answered 'yes' on unspecified chronic medical conditions in addition to people suffering from diabetes, hypertension, chronic allergies and chronic skin problems. Group V ($n=91$) consisted of people suffering from cancer. Approximately one third of the respondents, group VI ($n=631$), had indicated that they did not have any health problems.

Statistical analyses

People with health problems may have a tendency to be older than healthy people. Because HRQOL varies according to age and gender [7–9, 13, 14], it is not appropriate to compare directly the HRQOL of people with health problems against the reference values for the normal population. Instead, an adjustment for age and gender differences should be made. There are two principal approaches to this problem. First, it is possible to regard the age distribution of the reference population as being the standard to which all other data sets should be adjusted. This leads to standardised, or adjusted, mean scores for each of the disease groups. This is similar to the standardised rates which are commonly employed by epidemiologists when comparing incidence or prevalence rates among different subgroups. The second approach, which we illustrate in Appendix 2, is based upon the concept of obtaining 'expected' mean scores by using the population reference values and calculating the expected scores that would be observed for subjects of the same age and gender distribution as in the disease group. Since each disease group will have a different age and gender distribution, separate expected values are calculated for each group. The calculations make use of the basic reference data (Appendix 1) and are straightforward. We provide a detailed work example in Appendix 2. An alternative approach for obtaining expected scores is to use regression models to obtain individual patient expected values (Appendix 3).

Significance tests of the differences between observed and expected values (OBS − EXP) used 'bootstrap' methods to compensate for the non-normality of the data [26]. In this particular data set, allowance had to be made for overlap, since the people used for estimating HRQOL in patients were also included in the total sample and used to derive the

Table 1. *Response rate %* (n/N) *by age and gender*

Age group (years)	Male	Female	Total
18–29	63.9 (205/321)	70.9 (185/261)	67.0 (390/582)
30–39	66.1 (228/345)	73.3 (159/217)	68.9 (387/562)
40–49	65.7 (182/277)	67.4 (147/218)	66.5 (329/495)
50–59	73.2 (153/209)	70.0 (145/207)	71.6 (298/416)
60–69	73.0 (114/156)	75.9 (142/187)	74.6 (256/343)
≥70	66.3 (134/202)	58.6 (171/292)	61.7 (305/494)
Total	67.3 (1016/1510)	68.7 (949/1382)	67.9 (1965/2892)

population expected values; such correction will not be necessary when our reference data are used by others to compare against their own data sets.

The computation of the adjusted mean scores, calculation of P values and standard errors and the regression approach were carried out with STATA Statistical Software, release 5.0 [27], while version 6.1 of SPSS PC for Windows [28] was used for the rest of the statistical analyses. All calculations were performed after linearly transforming the mean scores of the EORTC QLQ-C30 to a 0–100 scale where high scores mean a high level of function (functional and QOL scales) or a high level of symptoms or problems (symptom scales and single items). So far, no general agreement exists of the definition of clinical significance. Nevertheless, differences of 10 or more on the 100 point scale might be regarded as such [13, 19, 29], as we use in this study.

RESULTS

Response rate

Of the 3000 packets mailed, 220 unanswered questionnaires were returned by the respondents themselves, their relatives or the postal authorities. The reasons for not answering were: incorrect address, 69; death, 2; disease, 17; incapability of completing, 20 (total 108). The remaining 112 persons had marked on the statement coupons that they did not want to participate. A total of 1965 people completed the questionnaires, yielding a response rate of 68% (1965/2892). Seventy-two per cent of the respondents returned the questionnaire before the reminder was sent.

The response rates were higher among women in four of six age groups (Table 1). The total response rate for men was 67.3% and for women 68.7%, varying from 58.5 to 75.9% for females and 63.9 to 73.2% for males by age.

Sample characteristics

The mean age of the sample was 47.4 years (range 19–93), consisting of 1016 men (52%) and 949 (48%) women. Sociodemographic characteristics are listed in Table 2. People suffering from cancer ($n=91$), were older than the rest, with a mean age of 65.9 years, while those who reported to be free of disease ($n=631$), were much younger (39.0 years). The 277 people with heart problems had a mean age of 55.6 years, while the mean ages in the other three groups (physical problems $n=694$, visual or hearing impairment $n=339$, chronic disease $n=779$) were 52.9, 55.6 and 52.4 years, respectively.

Mean scale and item scores

Sixty-eight per cent (1334) of the respondents answered 'yes' to one or more of the health questions. The most commonly reported health problems were: sciatica or chronic back problems ($n=523$), other chronic medical conditions ($n=330$), hypertension ($n=311$), paralysed/weak limb ($n=264$), chronic allergies ($n=202$), chronic obstructive lung problems ($n=154$) and congestive heart failure ($n=139$).

The scores on the different scales and items regarding HRQOL from these people are also included when estimating the reference values for the general population. Many health problems reported have a high prevalence in the general population and many people suffer from more than one of these conditions.

Table 3 shows the differences between the observed mean scores for each of the disease groups and the expected mean scores one would get, if the age and gender distributions were the same in the disease groups as in the total sample. Furthermore, the differences between observed and expected

Table 2. *Sociodemographic variables*

	Total population *n* (%)
Number of responding participants	1965
Mean age (years, range)	47.4 (19–93)
Gender	
Male	1016 (52)
Female	949 (48)
Social status	
Single, separated, divorced, widowed	920 (47)
Married/living with partner	1045 (53)
Living situation	
Living alone	329 (17)
Living with others (spouse/partner, parents/parents-in-law, children, other adults)	1626 (82)
Unknown	10
Education	
Elementary school up to 10 years	577 (29)
High school, A levels, vocational education, commercial college/general college	845 (43)
University/college	508 (26)
Unknown	35 (2)
Work	
Working in a paying job/self-employed	1120 (57)
Working at home, retired	418 (21)
Studying, military service, unemployed, receiving social welfare	376 (19)
Unknown	51 (3)
Type of work	
Worker, fisherman, farmer, driver	716 (36)
Subordinate staff, lower professional	639 (33)
Management position	333 (17)
Self-employed	98 (5)
Unknown	179 (9)

Table 3. Observed (OBS), age/gender equivalent expected (EXP) mean scores and deviations (OBS – EXP) for people reporting various past or present health problems

	Total sample	Cancer			People reporting no health problems			Cardiac problems (angina, myocardial infarction, chronic obstructive lung disease)			Physical diseases (arthritis, sciatica, weak/paralysed limb)			Visual or hearing impairment			Chronic disease (hypertension, chronic skin problems, diabetes, allergies, other chronic medical problems)		
	Male: 1016 Female: 949	Male: 34 Female: 57			Male: 353 Female: 278			Male: 145 Female: 132			Male: 329 Female: 365			Male: 169 Female: 170			Male: 380 Female: 399		
*n**	1965	91			631			277			694			339			779		
Mean age (years, range)	47.7 (19–93)	65.9 (30–92)			39.0 (19–86)			55.6 (19–93)			52.9 (19–92)			55.6 (19–93)			52.4 (19–93)		
		OBS	EXP	OBS – EXP	OBS	EXP	OBS – EXP	OBS	EXP	OBS – EXP	OBS	EXP	OBS – EXP	OBS	EXP	OBS – EXP	OBS	EXP	OBS – EXP
Functioning scales†																			
Cognitive	86.5	76.8	82.5	– 5.7	94.0	88.1	5.9	75.9	83.5	– 7.6	79.1	85.4	– 6.3	75.4	83.7	– 8.3	81.2	85.5	– 4.3
Emotional	82.8	77.9	83.4	– 5.5	89.6	82.5	7.1	74.0	83.7	– 9.7	76.8	82.8	– 6.0	77.0	83.7	– 6.7	77.1	82.9	– 5.8
Physical	89.9	72.0	79.8	– 7.8	98.9	93.1	5.8	73.3	83.9	– 10.6	79.6	86.9	– 7.3	78.6	83.3	– 4.7	82.5	87.3	– 4.8
Role	92.8	77.6	86.0	– 8.4	99.3	95.1	4.2	79.7	88.6	– 8.9	85.0	91.1	– 6.1	83.4	88.5	– 5.1	87.3	91.3	– 4.0
Role new	83.3	61.7	76.3	– 14.6	96.2	86.3	9.9	63.7	78.9	– 15.2	70.0	81.2	– 11.2	69.7	78.6	– 8.9	73.4	81.6	– 8.2
Social	85.8	70.4	82.4	– 12.0	95.8	87.4	8.4	68.7	83.8	– 15.1	76.1	84.7	– 8.6	76.4	83.9	– 7.5	76.4	84.8	– 8.4
Global QOL†	73.7	59.7	70.3	– 10.6	84.7	75.0	9.7	57.7	71.7	– 14.0	62.7	72.7	– 10.0	63.7	71.6	– 7.9	65.0	72.8	– 7.8
Global health/QOL†	75.3	59.9	71.6	– 11.7	86.6	76.8	9.8	58.0	73.2	– 15.2	64.2	74.3	– 10.1	65.4	73.2	– 7.8	66.1	74.4	– 8.3
Symptom scales‡																			
Fatigue	28.8	42.6	31.3	11.3	18.4	28.0	– 9.6	45.8	29.8	16.0	39.0	29.4	9.6	39.1	30.1	9.0	37.9	29.3	8.6
Nausea/vomiting	4.0	10.4	4.7	5.7	2.0	3.8	– 1.8	9.4	4.4	5.0	6.2	4.2	2.0	8.1	4.5	3.6	6.2	4.1	2.1
Pain	20.5	33.9	25.5	8.4	5.2	18.2	– 13	37.3	22.8	14.5	39.2	22.3	16.9	28.8	23.1	5.7	31.7	21.8	9.9
Single items‡																			
Appetite loss	7.5	12.7	8.8	3.9	2.9	7.4	– 4.5	18.1	8.3	9.8	11.6	7.5	4.1	13.5	8.5	5.0	11.2	7.6	3.6
Constipation	10.7	20.0	16.0	4.0	4.8	8.7	– 3.9	21.7	13.8	7.9	15.9	12.1	3.8	19.8	13.9	5.9	15.7	11.9	3.8
Diarrhoea	9.4	13.7	9.6	4.1	6.1	9.2	– 3.1	14.8	9.4	5.4	12.9	9.6	3.3	14.2	9.5	4.7	12.2	9.5	2.7
Dyspnoea	14.3	29.8	19.6	10.2	4.6	12.0	– 7.4	43.1	18.2	24.9	20.8	15.6	5.2	25.7	18.2	7.5	22.8	15.6	7.2
Financial impact	9.0	14.7	10.4	4.3	1.3	8.3	– 7.0	20.1	9.1	11.0	17.4	9.6	7.8	17.4	9.3	8.1	16.1	9.4	6.7
Sleep disturbance	20.4	36.4	26.3	10.1	10.0	18.2	– 8.2	34.0	22.7	11.3	29.9	22.3	7.6	30.4	23.6	6.8	28.6	21.8	6.8

**n* adds up to more than 1965 due to the overlap between groups. †Higher score indicates better function. ‡Higher score indicates more symptoms. QOL, quality of life; OBS, mean observed score for each disease group; EXP, expected mean scores one would get if the age and gender distributions were the same in the disease groups as in the total sample, all with standard errors between 0.4 and 0.6; OBS – EXP, difference between observed and expected mean scores (the differences in the disease groups had P values <0.005 or better, except in the cancer group with cognitive function, $P=0.008$; emotional function, $P=0.02$; appetite, $P=0.07$; constipation, $P=0.11$; diarrhoea, $P=0.08$ and financial difficulties, $P=0.09$).

(OBS–EXP) values are listed. The expected values are based upon the sample of 1965 and have standard errors of between 0.4 and 0.6 (the exact values depend upon both the scale or item and the age/gender distribution of the target group). All groups, except the cancer group, were so large that their OBS–EXP differences were significant with $P < 0.005$ or better. For cancer, most differences had P values < 0.005, the exceptions being: cognitive function, $P = 0.008$; emotional function, $P = 0.02$; appetite, $P = 0.07$; constipation, $P = 0.11$; diarrhoea, $P = 0.08$; and financial difficulties, $P = 0.09$.

The cancer patients showed great discrepancies (10 or more in difference) compared with the general population for all the functioning scales (but emotional), both QOL scales, fatigue and pain as well as constipation, dyspnoea and sleep disturbance. The observed mean values for the other disease groups also showed great deviations from the overall scores obtained in the total sample (first column) in the direction of lower HRQOL scores, more problems and symptoms as well as impaired functional levels. The cardiac patients was the group with the greatest deviations compared with the total sample, with a difference of 10 or more for 8 of 11 scales and 5 of 6 single items. For people suffering from visual and/or hearing impairments, values on the physical, cognitive, new role function and both QOL scales differed by 10 or more, as did fatigue sleep disturbance and dyspnoea. For people with physical impairments, the observed mean score for pain was most affected (39.2 versus 20.5) compared with the total population, with the new role function also demonstrating a considerable difference (70.0 versus 83.3). The same two scales also deviated most from the norm in the chronic disease group, 31.7 versus 20.5 and 73.4 versus 83.3, respectively.

For those who had reported no health problems, mean scores were higher on the functional and QOL scales and lower on the symptoms and single items when compared with the overall mean scores from the total sample. The difference was 10 or more for the social and the new role functioning scale, and for both QOL scales. Regarding the symptom scales, the greatest difference was found with pain, with 5.2 as the mean score in the 'healthy' group versus 20.5 in the total sample. Sleep disturbance and dyspnoea yielded a difference of 10 among the single items.

When comparing against the age and gender adjusted expected values, calculated as in Appendix 2, the major trends were still the same with lower mean scores on the functional and QOL scales, and higher scores on the single items and symptom scales for the disease groups. Emotional function was the only scale which was not affected for any of the disease groups, being equal or slightly better for all groups compared with the mean value from the total sample. The discrepancies between the observed values and the overall mean scores were strongly affected by the age and gender distribution of each disease group (Table 3).

DISCUSSION

This paper demonstrates that observed values on HRQOL in patient samples can be very misleading if age and gender are not taken into consideration. When allowance was made for these factors, the reduction in HRQOL scores in the disease groups compared with the values from the general population was smaller. We have presented how expected values for a given age- and gender-specific population can be computed by means of simple calculations or a regressional approach. Although this methodology is known from clinical trials, it is seldom used in research on HRQOL. The discrepancies between observed and expected values provide additional, clinically important information of the direct impact of age and gender on the scores from a HRQOL questionnaire, because comorbidities are taken into account.

For some purposes, it might be interesting to compare cancer patients against those reporting no health problems. However, cancer patients, like members of the general population, may suffer from other chronic diseases. Therefore, the most appropriate reference group will usually be the general population and that is the group that we selected for this comparison. The group of cancer patients in this study might have consisted of mainly cancer survivors. Although there were few subjects in this group ($n = 91$), they provide yet another reference group against which patients with newly diagnosed cancer, for example, might be compared.

Although we illustrate the calculations using the age-grouped data presented previously [13], a variation on this approach is to use the individual patient raw data and apply regression modelling to fit an equation which includes age and sex. This can be used to generate expected (predicted) values for each individual. In principle, this would be a slightly more accurate method, since it makes full use of the individual values and involves fitting a smooth curve. In practice, however, it usually makes negligible difference to the estimates and has two disadvantages: it requires access to the individual patient data and the regression models will often involve non-linear functions of age which are possibly different for each gender.

Even though the 14 questions on past or present illness are recommended when reporting QOL data from the SF-36 [7], few publications report results regarding this information. The SF-36 manual [7] has tabulated reference data for five diagnoses determined by physicians (hypertension, congestive heart failure, diabetes type II, acute myocardial infarction, clinical depression). All disease groups showed reduced functional status, except on the mental health scale. The greatest differences were found in people suffering from congestive heart failure. However, no adjustments were made for sociodemographic data or comorbidities. The Swedish study using the SF-36 [14], reported great reduced functioning on all scales for those who had stated that they were handicapped or suffering from long-standing chronic illness. Compared with population data for the same age groups, physical function was mostly impaired. Reduced functioning on all scales of the SF-36 (except mental health and emotional) was also found in the U.K. for patients suffering from chronic physical disease compared with population scores [8]. In this study, disease status was verified by a general practitioner. The observed values for each of the disease groups in our study were lower on the functional and QOL scales, meaning impaired function and reduced HRQOL, and higher on the symptom scales and single items, indicating more problems than those found in the general population [13]. The differences between the observed values and the calculated expected values based on the mean scores in the corresponding age and gender brackets in the general population, were large. This was especially true for cardiac patients, maybe attributable to the fact that the health problems included in this group are also frequent in the general population. Regarding the group of cancer patients in the sample, the expected values calculated were close to those of the reference population, strongly deviating from the observed mean scores. The age composition of the cancer patients was, therefore, important to consider. Another possible explanation is that the group might have consisted of mainly cancer survivors, for example, people who had terminated treatment in the past and who were probably free of disease. This is supported by the fact that the expected values in the group were closer than the observed to the scores obtained from the EORTC QLQ-C30 in a group of cancer out-patients clinically free of disease [18].

Although we used 'bootstrap' methods, because of the severe non-normality of much of the data, the results are extremely close to those obtained by carrying out *t*-tests. For most applications, a *t*-test will be adequate. Also, although the standard errors of the expected values vary, we suggest that in practice a standard error of 0.5 will be an adequate approximation. Investigators would need access to our full data set in order to calculate the exact standard error for the age/gender adjusted expected values.

One has to be cautious, however, when interpreting scores from people suffering from different diseases. First, the health problems were not verified by any clinical assessment and based on self-report only. Second, the respondents were collapsed into six categories according to health problems and third, many people suffer from more than one problem. Nevertheless, common health problems as asked for here are always prevalent in any Western population and might contribute to more accurate values from a general population in the construction of reference values.

What should be kept in mind is that the EORTC QLQ-C30 is a questionnaire which was developed and recommended for use among cancer patients and not intended for HRQOL assessments in the general population. It is important to notice that the intention with the reference values from the general population is to serve as comparison values with results from cancer patients, not to assess HRQOL in the general population *per se*. This is probably better assessed with generic HRQOL questionnaires. One might also question the relevance of the questions to the general population. However, the data give an estimate of the level of problems for the issues raised by the questionnaire. The simplicity of the layout and the uniformity of the answer categories enhance the ease of completing and it seemed acceptable to people in general, underlined by the high response rate of 68%. The ability of the questionnaire to discriminate between groups of people at a given time point and its clinical usefulness were supported by the distinct differences being demonstrated between subgroups of people with respect to gender, age and disease status.

We do not imply, however, that the EORTC QLQ-C30 should be used in non-cancer patients in general. Nor do we advocate that age and gender are the only two background factors regardless of comorbidities to consider when interpreting data on HRQOL. We think it is important for clinicians to get reference data on the EORTC QLQ-C30 and guidelines on how to use these. This has long been requested as an aid regarding counselling and treatment. Disease-specific modules [21–25] added to the core questionnaire will yield additional information about specific problems pertaining to certain cancer diagnoses. In non-randomised studies, the impact of age and gender on the scores observed has to be examined carefully and if necessary our approach can be used to compute adjusted values.

In conclusion, cancer patients often experience reduced HRQOL and report more problems than the general population. Nevertheless, not all of this is necessarily attributable to the cancer diagnosis or its treatment, but might, for example, be due to increasing comorbidities with increasing age. Data on HRQOL have to be interpreted with caution, due to the many confounding factors which might have a profound impact on HRQOL. As demonstrated in this paper, age and gender should always be taken into consideration, as well as concomitant diseases which are prevalent both in the general population and among cancer patients. Two simple and feasible methods are suggested on how to obtain more reliable estimates of the mean scores observed in patient studies.

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**, 365–376.
2. de Haes JCJM, van Knippenberg FCE, Neijt JP. Measuring psychological and physical distress in cancer patients: structure and application of the Rotterdam Symptom Checklist. *Br J Cancer* 1990, **62**, 1034–1038.
3. Schipper H, Clinch J, McMurray A, *et al.* Measuring the quality of life of cancer patients: the functional living index-cancer: development and validation. *J Clin Oncol* 1984, **2**, 472–483.
4. Cella DF, Tulsky DS, Gray G, *et al.* The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol* 1993, **11**, 570–579.
5. Schag CAC, Heinrich RL, Ganz PA. The Cancer Inventory of Problems Situations: an instrument for assessing cancer patients' rehabilitation needs. *J Psychosoc Oncol* 1990, **9**, 83–102.
6. Bergner M, Bobbitt R, Carter W, *et al.* The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981, **14**, 787–805.
7. Ware JE. *SF-36 Health Survey Manual and Interpretation Guide.* Boston, New England Medical Center, 1993.
8. Brazier JE, Harper R, Jones NMB, *et al.* Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *Br Med J* 1992, **305**, 160–164.
9. Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *Br Med J* 1993, **306**, 1437–1440.
10. Garratt AM, Ruta DA, Ruta M, *et al.* The SF-36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *Br Med J* 1993, **306**, 1440–1444.
11. Klee M, Groenvold M, Machin D. Quality of life of Danish women: population based norms for the EORTC QLQ-C30. *Qual Life Res* 1997, **6**, 27–34.
12. Schag CAC, Heinrich RL. *Cancer Rehabilitation Evaluation System (CARES). Manual*, First edition. Los Angeles, CARES Consultants, 1989.
13. Hjermstad MJ, Fayers PM, Bjordal K, *et al.* Health related quality of life in the Norwegian general population assessed by the European Organization for Research and Treatment of Cancer core quality of life questionnaire: the QLQ-C30 (+3). *J Clin Oncol* 1998, **16**, 1188–1196.
14. Sullivan M, Karlsson J, Ware JE. *Hälsoenkät SF-36, Svensk manual och tolkningsguide (SF-36 health survey). Swedish manual and interpretation guide.* Gothenburg, Sahlgrenska University Hospital, 1994.
15. Kaasa S, Bjordal K, Aaronson N, *et al.* The EORTC Core Quality of Life Questionnaire (QLQ-C30): validity and reliability when analysed with patients treated with palliative radiotherapy. *Eur J Cancer* 1995, **31A**, 2260–2263.
16. 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.
17. Ringdal GI, Ringdal K. Testing the EORTC Quality of Life Questionnaire on cancer patients with heterogeneous diagnoses. *Qual Life Res* 1993, **2**, 129–140.

18. Hjermstad MJ, Fosså DS, Bjordal K, *et al.* Test/retest study of the European Organization for Research and Treatment of Cancer core quality-of-life questionnaire. *J Clin Oncol* 1995, **13**, 1249–1254.
19. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res* 1996, **5**, 555–567.
20. Osoba D, Zee B, Pater J, *et al.* 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.
21. Sprangers MAG, te Velde A, Aaronson NK, *et al. EORTC Study Group on Quality of Life: the construction of a colorectal cancer module*. Amsterdam, Manchester, Edinburgh, Hamburg, EORTC Study Group on Quality of Life, 1992.
22. Bjordal K, Ahlner-Elmqvist M, Tollesson E, *et al.* Development of a European Organization for Research and Treatment of Cancer (EORTC) questionnaire module to be used in quality of life assessments in head and neck cancer patients. *Acta Oncol* 1994, **33**, 879–885.
23. Sprangers MAG, Groenvold M, Arraras S, *et al.* The EORTC breast cancer-specific quality of life questionnaire module (QLQ-BR23): first results from a three-country field study. *J Clin Oncol* 1996, **14**, 2756–2768.
24. Bergman B, Aaronson NK, Ahmedzai S, *et al.* The EORTC QLQ-LC13: a modular supplement to the EORTC core quality of life questionnaire (QLQ-C30) for use in lung cancer clinical trials. *Eur J Cancer* 1994, **30**, 635–642.
25. Blazeby J, Alderson D, Winstone K, *et al.* Development of an EORTC questionnaire module to be used in quality of life assessment for patients with oesophageal cancer. *Eur J Cancer* 1996, **32A**, 1912–1917.
26. Effron B, Tibshirani RJ. *An Introduction to Bootstrap*. New York, Chapman & Hall, 1993.
27. Stata Corporation. *STATA Statistical Software*, release 5.0. Texas, College Station, 1997.
28. Norusis MJ. SPSS for Windows, Base system User's Guide, release 6.1. Illinois, SPSS Inc, 1993.
29. Wisløff F, Eika S, Hippe E, *et al.* Measurement of health-related quality of life in multiple myeloma. *Br J Haematol* 1996, **92**, 604–613.

Acknowledgements—The authors wish to thank Anders Skjeggestad, regarding data management, and Gro Underland for good organising and practical help. Financial support by grant no. 93094/001 from the Norwegian Cancer Society.

APPENDIX 1

MEAN SCORES ON SCALES AND ITEMS FOR ALL RESPONDENTS BY GENDER AND AGE GROUPS

The following table was reproduced with permission from Hjermstad MJ, Fayers PM, Bjordal K, *et al.* *J Clin Oncol* 1998, **16**, 1188–1196.

	Male (age range, years)							Female (age range, years)							
	All (*n* = 1016)	18–29 (*n* = 205)	30–39 (*n* = 228)	40–49 (*n* = 182)	50–59 (*n* = 153)	60–69 (*n* = 114)	≥ 70 (*n* = 134)	All (*n* = 949)	18–29 (*n* = 185)	30–39 (*n* = 159)	40–49 (*n* = 147)	50–59 (*n* = 145)	60–69 (*n* = 142)	≥ 70 (*n* = 171)	Total (*n* = 1965)
Functioning scales*															
Cognitive	87.1	91.6	88.9	89.5	86.5	82.7	77.6	85.8	89.5	87.3	86.9	86.1	86.4	77.9	86.5
Emotional	85.4	87.8	83.7	84.6	83.9	84.8	87.7	79.9	78.7	77.2	76.8	83.2	79.4	84.4	82.8
Physical	93.2	98.0	97.3	91.1	92.9	89.1	77.7	86.4	97.6	93.6	89.9	87.6	78.2	67.7	89.9
Role	94.4	98.3	97.8	97.2	93.0	89.8	82.8	91.0	97.8	94.9	92.5	92.0	89.4	77.8	92.8
Role, new	85.7	92.4	91.0	85.5	84.4	79.0	73.3	80.6	88.4	85.9	82.6	81.3	75.5	68.1	83.3
Social	87.7	94.5	89.4	89.0	84.4	80.9	81.6	83.6	88.9	84.0	80.0	83.9	83.4	80.1	85.8
Global QOL*	75.4	78.5	77.5	76.0	73.5	73.6	70.2	71.7	76.1	73.4	71.2	72.7	69.4	66.4	73.7
Global health/QOL* new	77.3	80.1	78.8	79.0	75.9	75.2	71.4	73.2	77.9	75.7	72.7	74.1	70.2	67.6	75.3
Symptom scales†															
Fatigue	25.0	23.7	25.9	23.3	25.8	23.6	28.5	32.8	32.7	32.6	30.9	30.3	32.5	37.6	28.8
Nausea/vomiting	2.9	2.8	3.2	2.5	3.1	2.3	3.3	5.2	6.5	3.5	5.3	3.8	3.5	8.4	4.0
Pain	16.9	10.6	14.5	17.0	20.6	21.3	22.5	24.4	15.9	19.6	26.4	25.8	32.1	29.0	20.5
Single items†															
Appetite loss	5.7	8.4	5.4	3.9	3.2	4.2	8.6	9.4	11.9	10.2	8.5	3.7	5.9	15.3	7.5
Constipation	7.1	4.7	3.5	4.6	8.3	9.2	17.7	14.6	10.3	12.7	12.0	14.1	16.3	23.2	10.7
Diarrhoea	9.3	7.5	7.8	11.6	11.8	7.8	9.9	9.4	7.6	9.2	10.4	11.2	9.2	9.5	9.4
Dyspnoea	13.4	9.4	9.5	11.7	14.2	16.1	26.0	15.2	11.7	10.4	12.8	11.4	19.5	25.5	14.3
Financial impact	7.3	3.8	7.3	7.9	9.2	13.3	4.7	10.8	5.9	11.0	13.9	12.3	11.9	11.3	9.0
Sleep disturbance	15.5	11.7	14.5	15.2	18.7	18.6	17.3	25.6	16.6	21.2	24.8	24.8	33.1	35.4	20.4

*Higher score indicates better function. †Higher score indicates more symptoms.

APPENDIX 2

REFERENCE VALUES [13] FOR THE COGNITIVE FUNCTIONING SUBSCALE, BROKEN DOWN BY AGE GROUP AND GENDER

Age group (years)	Female	Male	Total
18–29	89.5	91.6	90.6
30–39	87.3	88.9	88.2
40–49	86.9	89.5	88.3
50–59	86.1	86.5	86.3
60–69	86.4	82.7	84.8
≥ 70	77.9	77.6	77.8
Total	85.8	87.1	86.5

AGE AND GENDER DISTRIBUTION OF THE SUBJECTS WITH CARDIAC PROBLEMS—NUMBER OF SUBJECTS IN EACH CATEGORY WHO COMPLETED THE COGNITIVE FUNCTIONING SUBSCALE

Age group (years)	Female	Male	Total
18–29	17	9	26
30–39	8	21	29
40–49	14	15	29
50–59	14	16	30
60–69	23	26	49
≥ 70	48	54	102
Total	124	141	265

From the reference values, we see that the 'expected' mean score for each of the 17 females aged 18–29 years is 89.5. Therefore, for 124 females with the same age distribution as the subjects with cardiac problems, we would expect a total sum score of:

$$17\times89.5+8\times87.3+14\times86.9+14\times86.1 \\ +23\times86.4+48\times77.9=10\,368.3$$

Therefore, the expected mean value for females is

$$10\,368.3/124= \quad 83.6$$

Similarly the expected sum score for the males is:

$$9\times91.6+21\times88.9+15\times89.5+ \\ 16\times86.5+26\times82.7+54\times77.6=11\,758.4$$

Therefore, the expected mean value for males is

$$11\,758.4/141= \quad 83.4$$

Consequently the expected mean value for the 265 subjects is

$$22\,126.7/265= \quad 83.5$$

Since the observed mean score for cognitive function in patients with cardiac problems was 75.9, they have a mean score which is 75.9 − 83.5 = 7.6 below the expected value for people of similar age and gender.

Note that the overall mean score in the population was 86.5, and that if we used that number without making adjustment for age and gender distributions we would have estimated the subjects with cardiac problems to be 10.6 below the reference value.

APPENDIX 3

Using the regression approach with the individual subject reference data, we start by fitting a suitable function of age and gender. Since inspection of the plotted data suggests a non-linear age relationship, a 'dummy variable' is created for a quadratic age^2 term. An age/gender interaction term was also put into the model, allowing different age trends for the two genders.

```
agegen=age*(gender-1)
agesq=age*age/100
```

We regress `xcf` (cognitive function, mean score) against age, agesq, agegen, gender
Using the statistical package STATA [26], we obtain:

```
   Source |       SS       df       MS                  Number of obs =    1926
----------+------------------------------               F(  4,  1921) =   23.74
    Model |   35035.892     4  8758.97299               Prob > F      =  0.0000
 Residual |  708842.036  1921  368.996375               R-squared     =  0.0471
----------+------------------------------               Adj R-squared =  0.0451
    Total |  743877.928  1925  386.430092               Root MSE      =  19.209

------------------------------------------------------------------------------
      xcf |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
----------+-------------------------------------------------------------------
      age |   .2703326   .1425383      1.897   0.058      -.0092134    .5498786
    agesq |  -.4462821   .1359431     -3.283   0.001      -.7128937   -.1796706
   agegen |  -.0852684   .0487837     -1.748   0.081      -.1809429    .0104062
   gender |   4.431058   2.453924      1.806   0.071      -.3815777    9.243693
    _cons |   80.37667   5.078471     15.827   0.000       70.41678    90.33657
------------------------------------------------------------------
```

This regression equation, which allows for age and gender, can then be used to predict the most likely values for each person with cardiac problems.

With STATA [26], this is done by the simple command

```
predict expCF
```

The mean value of the expected cognitive function scores for the 265 subjects with cardiac problems was 83.3. This is almost the same as the estimate using the grouped data presented in Ref. 13.