PERGAMON

European Journal of Cancer 36 (2000) 1796–1807

European Journal of Cancer

www.ejconline.com

A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (EORTC QLQ-H&N35) in head and neck patients

K. Bjordal [a,*], A. de Graeff [b], P.M. Fayers [c], E. Hammerlid [d], C. van Pottelsberghe [e], D. Curran [e], M. Ahlner-Elmqvist [f], E.J. Maher [g], J.W. Meyza [h], A. Brédart [i], A.L. Söderholm [j], J.J. Arraras [k], J.S. Feine [l], H. Abendstein [c], R.P. Morton [m], T. Pignon [n], P. Huguenin [o], A. Bottomly [e], S. Kaasa [c] on behalf of the EORTC Quality of Life Group

[a] *The Department of Radiotherapy and Oncology, The Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway*
[b] *University Medical Center Utrecht, The Netherlands*
[c] *University Hospital, Trondheim, Norway*
[d] *Sahlgrenska University Hospital, Goteborg, Sweden*
[e] *EORTC Data Center, Brussels, Belgium*
[f] *University Hospital, Malmo, Sweden*
[g] *Mount Vernon Hospital, Northwood, UK*
[h] *Maria Sklodowska-Curie Memorial Cancer Centre, Warsaw, Poland*
[i] *Institut Jules Bordet, Brussels, Belgium*
[j] *Department of Maxillofacial Surgery, Helsinki University Central Hospital, Helsinki, Finland*
[k] *Hospital of Navarre, Pamplona, Spain*
[l] *McGill University Montreal, Canada*
[m] *Green Lane Hospital, Auckland, New Zealand*
[n] *Hopital d'adultes de la Timone, Marseille, France*
[o] *Institut fur Radioonkologie, Universitatsspital, Zurich, Switzerland*

Received 5 July 1999; received in revised form 28 March 2000; accepted 21 June 2000

Abstract

This study tests the reliability and validity of the European Organization for Research and Treatment of Cancer (EORTC) head and neck cancer module (QLQ-H&N35) and version 3.0 of the EORTC Core Questionnaire (QLQ-C30) in 622 head and neck cancer patients from 12 countries. The patients completed the QLQ-C30, the QLQ-H&N35 and a debriefing questionnaire before antineoplastic treatment or at a follow-up. 232 patients receiving treatment completed a second questionnaire after treatment. Compliance was high and the questionnaire was well accepted by the patients. Multitrait scaling analysis confirmed the proposed scale structure of the QLQ-H&N35. The QLQ-H&N35 was responsive to differences between disease status, site and patients with different Karnofsky performance status, and to changes over time. The new physical functioning scale (with a four-point response format) of version 3.0 of the QLQ-C30 was shown to be more reliable than previous versions. Thus, the QLQ-H&N35, in conjunction with the QLQ-C30, appears to be reliable, valid and applicable to broad multicultural samples of head and neck cancer patients.

Keywords: Quality of life; Head and neck cancer; Assessment; Questionnaire; Validation; EORTC QLQ-C30; EORTC QLQ-H&N35

1. Introduction

Having head and neck (H&N) cancer may be a shattering experience. These patients not only have to face a life-threatening disease, but also have to deal with the impact of the disease and its treatment on appearance

* Corresponding author. Tel. +47-22-93-40-00; fax +47-22-93-59-42.
E-mail address: k-bjorda@online.no (K. Bjordal).

0959-8049/00/$ - see front matter
PII: S0959-8049(00)00186-6

and on important functions like eating, swallowing, breathing and communication [1]. Treatment strategies are aimed not only at increasing the chances of cure but also at maintaining health-related quality of life (HQOL), for example, preservation of speech [2]. Measuring HQOL in these patients is therefore of great importance.

The European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Group has developed a strategy for measuring HQOL in clinical trials [3]. A tumour-specific module, e.g. a breast cancer [4] or a lung cancer module [5], is used in conjunction with a general cancer questionnaire — the QLQ-C30 [6]. These modules are developed according to the guidelines of the EORTC Quality of Life Group [7].

An EORTC H&N cancer-specific module was developed according to these guidelines [8]. The current 35-item version (QLQ-H&N35) (see the Appendix) has been translated into 10 languages following the EORTC Quality of Life Study Group translation guidelines [9].

A preliminary reliability and validity study of the module which started in 1993 included 500 patients with newly diagnosed disease in four centres in Norway, Sweden and The Netherlands [10]. Following this study, a structure consisting of seven symptom scales (pain, swallowing, senses, speech, social eating, social contact and sexuality), six symptom items (problems with teeth, problems with opening mouth, dry mouth, sticky saliva, coughing and feeling ill) and five additional dichotomous items (related to the use of painkillers, nutritional supplements and feeding tube, and to weight decrease or weight increase) was proposed (see the Appendix).

The final EORTC field study of the module was initiated in 1995. The primary aim of this study was to test the postulated scale structure of the QLQ-H&N35 with regard to its reliability and validity in a large international sample of patients, including patients with newly diagnosed disease, recurrent disease and disease-free patients. The secondary aim of the study was to test the reliability and validity of version 3.0 of the QLQ-C30 (with a changed response format of the physical functioning scale) in this patient group.

The results from the preliminary validation study indicated that some of the scales (senses and speech) performed better in some patient subgroups than in others. A few items (trouble eating and painful throat) could be included in more than one scale, and other items (problems with teeth and the last five items) were candidates for removal. This paper explores these issues, with a special emphasis on the clinical validity of scales and single items. This is the definitive report about the psychometric properties of the QLQ-H&N35, and a copy of the module is shown in the Appendix.

2. Patients and methods

2.1. Patients

Patients were eligible for the study if they had newly diagnosed or recurrent squamous cell carcinoma of the larynx, oral cavity, oro-, naso- or hypopharynx undergoing active treatment (group I) or were disease-free 1–3 years after treatment (group II).

Group I was subdivided into five subgroups: newly diagnosed patients with laryngeal cancer receiving radiotherapy as first treatment (IA), or with cancer of the oral cavity or pharynx receiving radiotherapy (IB) or surgery (IC) as first treatment, and patients with recurrent disease (independent of site or previous treatment) receiving surgery as salvage treatment (ID) or chemotherapy (IE).

Group II was subdivided into two subgroups, independent of previous treatment: disease-free patients with cancer of the larynx (IIA) or cancer of the oral cavity or pharynx (IIB).

Exclusion criteria were: expected survival less than three months; inability to understand the questionnaire; cognitive and/or mental impairment; brain metastases or intracranial extension of the tumour with cognitive impairment; other previous or concurrent malignancies; participation in another HQOL study interfering with the field study. There was no limit on age or performance status. Informed consent (oral or written) was obtained from all patients and the study was approved by the national or institutional ethical committees.

2.2. Questionnaires and data collection

Patients completed the EORTC QLQ-C30 (version 3.0), the EORTC QLQ-H&N35 and a debriefing questionnaire before the start of treatment (group I) or at a regular follow-up visit 1–3.5 years after finishing active treatment (group II). Patients from group I completed a second set of questionnaires within 7 days before or after completion of radiotherapy (groups IA and IB), 3.5–5.5 weeks after surgery (groups IC and ID), or up to 4 days before the third cycle of chemotherapy or at 6 weeks in case of weekly or continuous chemotherapy (group IE). Patients were considered to be evaluable if they had completed at least one questionnaire. The EORTC QLQ-C30 [6] is a cancer-specific questionnaire which has been used in H&N cancer patients [10–17]. The current version 3.0 differs in three respects from the first version: the role functioning and overall QOL scales have been changed and the dichotomous response format of the items of the physical functioning scale has been replaced by four-point Likert-type response categories.

The EORTC QLQ-H&N35 [8,10] is meant to be used in conjunction with the QLQ-C30 in patients with H&N cancer, irrespective of site, stage and treatment. It contains both single items and scales. The time frame

which the module addresses is 'during the last week'. The first 30 items are scored on a four-point Likert scale ('not at all', 'a little', 'quite a bit' and 'very much'), whereas the last five items have a no/yes format.

The scores of the QLQ-C30 and of the QLQ-H&N35 are transformed to a scale of 0–100, with a high score implying a high level of symptoms or problems (both questionnaires), or a high level of functioning or global QOL (QLQ-C30).

The debriefing questionnaire contained questions about the time required to complete both questionnaires, the need for assistance and the presence of questionnaire items which were confusing, difficult to answer or upsetting.

The following sociodemographic and clinical data were collected: gender, age, marital status, cohabitation, education, employment, site of the tumour or relapse, stage, previous and/or subsequent treatment, Karnofsky performance status [18], weight loss, selected symptoms/side-effects (graded according to the National Cancer Institute (NCI) toxicity criteria [19], co-morbidity and the use of pain medication or antiemetics.

2.3. Analysis

2.3.1. Reliability

The reliability (internal consistency) of the scales of the QLQ-C30 and QLQ-H&N35 was assessed using Cronbach's alpha coefficient. A value of greater than 0.70 was considered to be adequate [20].

2.3.2. Construct validity

Construct validity refers to whether the hypothesised scales are actually measuring the underlying construct. Multitrait scaling analysis [21] was used to confirm the proposed scales of the QLQ-H&N35 [10]. Evidence of item convergent validity was defined as a correlation of 0.40 or greater between an item and its own scale (corrected for overlap). Item discriminant validity was based on the comparison of the correlation with its hypothesised scale compared with other scales. If the correlation between an item and its own scale was more than two standard errors below its correlation with another scale a 'definite' scaling error was counted. If the correlation was within two standard errors, a 'probable' scaling error was counted. Where necessary, items were moved to other scales as long as the new scales were clinically sensible. Scaling analysis was performed separately for each assessment point and for the different subgroups.

Correlations between scales were explored in order to evaluate the discriminant validity. A consistently high correlation (>0.70) may indicate that two scales assess the same or highly related constructs.

2.3.3. Criterion validity

Criterion validity refers to the ability of the questionnaire to measure what it should measure. In the absence of a 'gold standard', the method of known groups comparison [22] was used to assess criterion validity. For this purpose, the ability of both questionnaires to discriminate between subgroups of patients differing in terms of clinical status (such as site, stage, performance status, weight loss and symptom/toxicity rating) was assessed.

2.3.4. Responsiveness to change over time

An important aspect of the clinical validity of a questionnaire is its ability to detect change over time. A difference of 10 units on a 0–100 scale has been suggested to be clinically important for the QLQ-C30 [23]. Change over time was assessed in group I patients comparing pretreatment with post-treatment values.

2.3.5. Statistical methods

The statistical software SPSS 6.1 for Windows [24] was used for the descriptive analysis. The Kruskal–Wallis test was used to detect differences between groups. The relative efficiency (RE) is the relative ability of the scales to detect group differences [25]. The scale or item with the lowest Chi-squared statistic in the Kruskal–Wallis test was assigned an RE-value of 1; the RE-values of the other scales and items were calculated as the ratio of their Chi-squared statistic over that of the reference item or scale. The Mann–Whitney test was used to detect change of mean scores over time. The standardised response mean (SRM) was used to assess the size of the change relative to the background variability in the measurements. The SRM was calculated as the difference of the mean scores over time divided by the standard deviation of these changes in scores. A SRM of 0.20, 0.50 and 0.80 is considered small, medium and large, respectively [26]. One-way ANOVA was used to compare means of scales by symptom/toxicity grading. To protect against problems associated with multiple significance testing, a conservative P value of <0.01 was regarded as statistically significant. The MAP-R [21] and STATA-software [27] were used for the reliability and scaling analyses.

3. Results

3.1. Patient characteristics

A total of 622 eligible and evaluable patients was included in the study: 204 patients with newly diagnosed disease (groups IA, IB and IC), 58 patients with recurrent disease (groups ID and IE) and 360 disease-free patients (group II) (Table 1). This patient group from 12 countries comprising nine languages represented the full range of sites, stages and treatments of H&N cancer.

Table 1
Patient characteristics ($n = 622$)

	Active treatment					Disease-free		
	Newly diagnosed			Recurrent				
	Larynx	Oral/pharynx		All sites		Larynx	Oral/pharynx	
	Radio-therapy	Radio-therapy	Surgery	Surgery	Chemotherapy			
	IA	IB	IC	ID	IE	IIA	IIB	Total
	n (%) $n = 88$	*n* (%) $n = 68$	*n* (%) $n = 48$	*n* (%) $n = 32$	*n* (%) $n = 26$	*n* (%) $n = 185$	*n* (%) $n = 175$	*n* (%) $n = 622$
Male/female	81 (92)/ 7 (8)	54 (79)/ 14 (21)	29 (60)/ 19 (40)	24 (75)/ 8 (25)	22 (85)/ 4 (15)	164 (89)/ 21 (11)	131 (75)/ 44 (25)	505 (81)/ 117 (19)
Median age (years) (range)	66 (47–86)	65 (38–89)	66 (32–87)	62 (31–84)	55 (31–83)	65 (23–89)	60 (22–91)	63 (22–91)
Site								
Oral cavity	–	20 (29)	43 (90)	12 (38)	9 (35)	–	107 (61)	191 (31)
Oropharynx	–	31 (46)	3 (6)	3 (9)	5 (19)	–	41 (23)	83 (13)
Nasopharynx	–	9 (13)	–	1 (3)	3 (12)	–	8 (5)	21 (3)
Hypopharynx	–	7 (10)	2 (4)	2 (6)	2 (8)	–	19 (11)	32 (5)
Larynx	88 (100)	–	–	14 (44)	6 (23)	185 (100)	–	293 (47)
Unclassified	–	1 (1)	–	–	1 (4)	–	–	2 (0.3)
AJCC stage								
I	48 (55)	7 (10)	12 (25)	–	–	–	–	67 (33)
II	16 (18)	16 (24)	11 (23)	–	–	–	–	43 (21)
III	14 (16)	22 (32)	10 (21)	–	–	–	–	46 (23)
IV	10 (11)	23 (34)	15 (31)	–	–	–	–	48 (24)
Previous treatment[a]								
Surgery	–	–	–	20 (63)	16 (62)	68 (37)	119 (68)	223 (53)
Radiotherapy	–	–	–	18 (56)	22 (85)	177 (96)	132 (75)	349 (83)
Chemotherapy	–	–	–	4 (13)	9 (35)	9 (5)	28 (16)	50 (12)
Karnofsky Performance status								
100	44 (50)	19 (28)	23 (48)	8 (25)	4 (15)	81 (44)	78 (45)	257 (41)
90	24 (27)	25 (37)	13 (27)	11 (34)	9 (35)	48 (26)	51 (29)	181 (29)
40–80	20 (23)	24 (35)	12 (25)	13 (41)	13 (50)	56 (30)	46 (26)	184 (30)
Countries								
Norway	8 (9)	9 (13)	–	1 (3)	2 (8)	8 (4)	8 (5)	36 (6)
Sweden	10 (11)	5 (7)	11 (23)	4 (13)	3 (12)	14 (8)	16 (9)	63 (10)
Finland	4 (5)	7 (10)	11 (23)	4 (13)	–	16 (9)	21 (12)	63 (10)
UK	21 (24)	19 (28)	2 (4)	3 (9)	–	29 (16)	18 (10)	92 (15)
The Netherlands	12 (14)	4 (6)	11 (23)	6 (19)	6 (23)	16 (9)	17 (10)	72 (12)
Belgium (Flem./French)	1 (1)/1 (1)	–	1 (2)/2 (4)	1 (3)/3 (9)	0/1 (4)	20 (11)/9 (5)	13 (7)/15 (9)	36/31 (11)
France	3 (3)	4	–	–	3 (12)	3 (2)	8 (5)	21 (3)
Switzerland (German)	–	–	–	–	–	11 (6)	6 (3)	17 (3)
Spain	10 (11)	6 (9)	1 (2)	4 (13)	9 (35)	17 (9)	17 (10)	64 (10)
Poland	11 (13)	6 (9)	2 (4)	2 (6)	2 (8)	15 (8)	13 (7)	51 (8)
Canada (English/French)	3 (3)/2 (2)	4 (6)/2 (3)	3 (6)/2 (4)	4 (13)/0	–	9 (5)/6 (3)	7 (4)/7 (4)	30/19 (8)
New Zealand	2 (2)	2 (3)	2 (4)	–	–	12 (6)	9 (5)	27 (4)

[a] Patients may have received more than one treatment.

3.2. Compliance

The percentage of missing items in both questionnaires was generally low (< 3%), with the exception of some items of the QLQ-H&N35: problems with teeth (3.0% missing), sexuality (10.5% and 13.4%, respectively) and weight gain (3.9%).

All patients in the study completed at least one questionnaire. Of the 262 patients of group I, only 30 (11%) failed to complete a second questionnaire, reasons

being: the patient felt too ill ($n=3$); the clinician or nurse felt the patient was too ill ($n=5$); administrative failure ($n=6$); patient withdrawal, no reason given ($n=4$); death ($n=5$); or other reasons ($n=7$).

A debriefing questionnaire was available for 529 patients (85%). It took the patients a median of 13 min to complete both questionnaires. 591 (95%) completed both questionnaires within 20 min. 162 (26%) of the patients received help to complete it, mostly by a psychologist, nurse or partner. The help usually consisted of reading the questions to the patient and completing the form for them. 68 (11%) found one or more items of the QLQ-H&N35 to be confusing or difficult to answer of whom half found the items on physical contact and sexuality to be difficult. 19 (3%) found items to be upsetting; in almost all cases, this related to the sexuality items.

3.3. Scale construction

Scaling analysis mostly confirmed the proposed scale structure of the QLQ-H&N35. The most problematic item was painful throat (hn4; part of the pain scale). For this item, definite scaling errors were observed for the total sample and for group I, and probable scaling errors for group II and for all sites (oral cavity, pharynx, and larynx), with hn4 showing the highest correlation with the swallowing scale.

Probable scaling errors were counted for the item about bothersome appearance (hn18). Although this item was hypothesised to be part of the social contacts scale, it was more highly correlated with the social eating scale than with its own scale. Similarly, the item about trouble eating (hn19), hypothesised to be part of the social eating scale, showed a higher correlation with the swallowing scale than with its own scale. When these analyses were repeated by group and by site (oral cavity, pharynx and larynx), probable scaling errors were quite consistently found for these two items. In the subgroup analyses, no other items were shown to have a scaling error, with the exception of item hn16 (being hoarse), which in the oral cavity subgroup was more highly correlated with the swallowing scale than with the speech scale.

3.4. Reliability

The internal consistency of the proposed scales was found to be adequate, with Cronbach's alpha coefficients of 0.81 (pain), 0.82 (swallowing), 0.72 (senses), 0.74 (speech), 0.87 (social eating), 0.83 (social contact), and 0.95 (sexuality). When the analysis was repeated for each group and site, all scales again showed Cronbach's alpha coefficients of >0.70, with the exception of the senses scale for which the Cronbach's alpha coefficient was 0.68 in the pharynx subgroup.

3.5. Distribution of scores

For both questionnaires, the distributions of scores were invariably skewed towards the positive range of responses (high values for functional scales and global QOL, and low values for symptom scales and single items) for all scales and single items. After treatment (group I), the distribution moved towards the negative side, but remained skewed. However, the full range of scores was observed for all scales and single items.

3.6. Construct validity of the QLQ-H&N35

Most of the scales of the QLQ-H&N35 showed a correlation of <0.70 with the scales of the QLQ-C30 or the other scales of the QLQ-H&N35, implying a high discriminant validity of the H&N cancer specific scales, with the exception of the swallowing and social eating scales, which showed a correlation of 0.73. The correlation between the pain scales of the QLQ-C30 and the QLQ-H&N35 was 0.63, and the correlation between the social functioning scale (scored as a functioning scale) of the QLQ-C30 and the social contact scale (scored as a symptom scale) of the QLQ-H&N35 was -0.58. None of the single items of the QLQ-H&N35 showed a correlation of >0.70 with any of the scales of either questionnaire.

3.7. Reliability and validity of the physical functioning scale of the QLQ-C30

A formal test of the new response format of the physical functioning scale of the QLQ-C30 has not yet been published. Table 2 compares the new version of this scale in newly diagnosed patients from this study to the version of the scale as used in the study of Bjordal and colleagues [10]. Baseline patient characteristics, including the mean Karnofsky score, were comparable in both studies. The Cronbach's alpha coefficient for the newer physical functioning scale, using four-point items, was 0.84 in contrast to the weaker alpha coefficient of 0.66 in the earlier scale. In both studies, very few patients (less than 4%) needed help with eating, dressing or using the toilet, and so in this group of patients item 5 contributed little to the overall scale; the other items correlated with the new scale more strongly than for the old scale. The new scale had a stronger association with the Karnofsky status.

3.8. Known groups comparisons

Differences in the scales and single items of both questionnaires between patients with newly diagnosed disease, patients with recurrent disease and disease-free patients are shown in Table 3. For the QLQ-C30, there were statistically and clinically significant differences for role, emotional and social functioning, and for fatigue,

Table 2
Comparison of the new (version 3.0) and old (version 1.0) physical functioning (PF) scales of the QLQ-C30 in newly diagnosed patients in the current study and in the study of Bjordal and colleagues [10], respectively[a]

	QLQ-C30 v3.0 (new PF-scale)	QLQ-C30 v1.0 (original PF-scale)
Number of patients	197	402
Median age (range) (years)	66 (32–89)	65 (31–88)[b]
Mean Karnofsky (S.D.)	89.9 (11.6)	89.9 (10.3)
Mean score PF (S.D.)	84.1 (21.4)	84.7 (21.2)
Cronbach's alpha coefficient	0.84	0.66
Item (1–5)-scale correlations	0.73, 0.79, 0.76, 0.68, 0.42	0.57, 0.58, 0.47, 0.40, 0.24
Ordered logistic regression, Karnofsky score on PF (Chi-square values)	55.9	45.9
Spearman's correlation, PF with Karnofsky score	0.42	0.29

S.D., standard deviation.
[a] 7 patients had missing values in the PF scale and were excluded in this analysis.
[b] To ensure comparability of patients across the studies, 13 patients younger than 30 years of age were excluded from analysis.

pain, insomnia, appetite loss and general QOL. Patients with recurrent disease had the worst values. For the QLQ-H&N35, there were significant differences for all the scales and single items (with the exception of teeth and coughing). Again, patients with recurrent disease had the worst values, with the exception of dry mouth. For the senses scale and the dry mouth, opening mouth, teeth and sticky saliva items, as well as social eating and contact items, disease-free patients had worse scores than patients with newly diagnosed disease. For the

Table 3
Differences in mean scores (± S.D.)[a] of scales and single items of the QLQ-C30 and the QLQ-H&N35 by disease status

	Newly diagnosed ($n=204$)	Recurrent ($n=58$)	Disease-free ($n=360$)	P value[b]	RE
QLQ-C30					
Physical functioning	84±21.4	79±21.1	85±18.8	0.2	21
Role functioning	81±30.2	63±37.3	84±25.4	<0.001	110
Emotional functioning	72±25.2	68±26.8	81±22.8	<0.001	148
Cognitive functioning	83±21.5	79±25.7	86±19.8	0.1	25
Social functioning	85±21.5	70±32.1	86±22.8	<0.001	95
Fatigue	26±24.7	36±29.4	21±23.6	<0.001	93
Nausea/vomiting	5±12.0	6±14.2	5±13.3	0.6	6
Pain	20±24.9	38±33.2	15±23.0	<0.001	208
Dyspnoea	20±27.8	16±25.1	20±29.5	0.6	6
Insomnia	28±31.6	33±35.3	22±30.5	0.006	57
Appetite loss	15±28.1	37±37.1	13±25.8	<0.001	19
Constipation	13±25.1	19±31.9	11±23.6	0.1	23
Diarrhoea	5±15.6	5±13.7	5±16.0	0.9	1
Financial problems	12±22.6	17±31.6	14±27.6	0.8	3
General QOL	63±23.9	55±22.8	73±21.7	<0.001	230
QLQ-H&N35					
Pain	21±22.5	34±29.9	13±18.0	<0.001	24
Swallowing	15±22.8	32±31.3	15±22.2	<0.001	105
Senses	8±18.4	26±29.8	19±28.9	<0.001	183
Speech	23±25.6	31±25.6	19±23.4	<0.001	94
Social eating	12±19.4	34±32.4	16±26.4	<0.001	165
Social contact	6±11.9	18±22.3	8±16.9	<0.001	88
Sexuality	27±34.0	46±38.8	25±33.1	<0.001	95
Teeth	17±27.4	19±33.3	19±30.8	0.9	1
Opening mouth	11±24.4	33±39.8	14±27.0	<0.001	118
Dry mouth	24±28.4	42±38.1	45±38.2	<0.001	237
Sticky saliva	20±26.8	45±38.9	37±37.2	<0.001	181
Coughing	27±28.2	24±30.7	24±28.4	0.3	12
Feeling ill	18±27.4	35±35.4	12±23.1	<0.001	170

RE, relative efficiency; QOL, quality of life.
[a] A high score for a functional scale or global QOL implies a high level of functioning or global QOL, whereas a high score for a symptom scale or single item implies a high level of symptoms.
[b] Kruskal–Wallis test.

QLQ-H&N35, the P values were generally $P < 0.001$, implying a high ability to distinguish between groups. The most sensitive scales and items (highest RE) were swallowing, senses, social eating, opening mouth, dry mouth, sticky saliva, and feeling ill. The least sensitive scale was the teeth item, with the baseline RE of 1.

With regard to differences between sites, for the QLQ-C30 there were significant differences only for the pain scale (21, 24 and 14, for oral cavity, pharynx and larynx, respectively) and the dyspnoea (14, 18 and 24, respectively) and appetite loss items (18, 24 and 10, respectively). $P \leqslant 0.01$ for those with significant differences ($P = 0.002$ for pain and 0.001 for dyspnoea). Differences between sites for the scales and single items of the QLQ-H&N35 are shown in Table 4. In addition, patients with cancer of the oral cavity or pharynx had worse scores for pain, swallowing, social eating, teeth problems and opening mouth, whereas patients with laryngeal cancer had worse scores for speech and coughing. Patients with pharyngeal cancer scored worst for the senses scale and the dry mouth and sticky saliva items. The least sensitive symptom single item and scales of the QLQ-H&N35 were feeling ill, social contact and sexuality which was set as baseline of 1.

Both the QLQ-C30 and the QLQ-H&N35 picked up few differences between stages in patients with newly diagnosed disease (data not shown). Statistically significant differences were seen only for the physical functioning and pain scales of the QLQ-C30, and the swallowing and social eating scales and the opening mouth item of the QLQ-H&N35. When data from the patients with laryngeal cancer and patients with oral cavity/pharyngeal cancer were analysed separately, the number of patients in each group was smaller and no significant differences between stages were found.

The differences in mean scale scores between patients with different Karnofsky performance status are shown in Table 5. Patients with a higher performance status consistently showed higher values for the functional and global QOL scales and lower values for the symptom scales and items.

When physician-rated symptom/toxicity scores for pain, mouth dryness, dysphagia, nausea/vomiting, diarrhoea, dyspnoea and speech were compared with the scores of the corresponding symptom scale or item, there was a high correlation, patients with higher grades reporting significantly higher scores (Table 6). The data for the second questionnaire showed the same pattern (data not shown).

3.9. Change over time

Change over time (i.e. comparison of scores before and after treatment in group I) is shown in Table 7. There was a significant deterioration of almost all functional and symptom scales and items of both questionnaires. The changes in the scores of the QLQ-C30 were generally of a lower order (mostly between 5 and 10, indicating a small to medium change, based on the SRM-values) than those of the QLQ-H&N35 (mostly 10–20, indicating a medium to large change, based on the SRM-values).

4. Discussion

The EORTC QLQ-H&N35 has been tested in two large series of patients. The first analysis was performed in a sample of 500 patients with newly diagnosed H&N cancer from four institutions in Norway, Sweden and The Netherlands [10]. The scale structure proposed in

Table 4
Differences in mean scores (±S.D.)[a] of scales and single items of the QLQ-H&N35 between sites (baseline questionnaires only; patients undergoing active treatment and disease-free patients combined) ($n = 620$)[b]

QLQ-H&N35	Oral cavity ($n = 191$)	Pharynx ($n = 136$)	Larynx ($n = 293$)	P value[c]	RE
Pain	24±23.1	26±25.9	10±14.7	< 0.001	180
Swallowing	19±23.7	26±27.4	10±20.1	< 0.001	147
Senses	14±23.5	23±30.6	14±25.6	0.002	33
Speech	16±21.4	17±21.0	28±26.8	< 0.001	81
Social eating	20±27.6	26±29.7	10±20.1	< 0.001	132
Social contact	10±18.0	10±17.5	7±14.8	0.6	3
Sexuality	28±36.4	29±34.4	26±33.1	0.8	1
Teeth	21±32.2	25±33.9	13±25.5	< 0.001	38
Opening mouth	22±32.6	24±34.0	5±16.4	< 0.001	156
Dry mouth	37±36.5	54±39.7	31±32.9	< 0.001	79
Sticky saliva	31±35.6	45±40.6	27±31.0	< 0.001	44
Coughing	19±28.5	26±29.1	28±27.9	< 0.001	47
Feeling ill	17±26.5	19±28.8	14±25.5	0.2	8

RE, relative efficiency.
[a] A high score implies a high level of symptoms.
[b] 2 patients were unclassified.
[b] Kruskal–Wallis test.

that first study has now been tested and validated in this larger and more diverse sample (including patients with recurrent disease and disease-free patients) from 12 countries. This makes the EORTC QLQ-H&N35 one of the most widely tested disease-specific HQOL modules in cancer patients. We have not yet explored cross-cultural differences. This will be the subject of a separate paper.

Table 5
Differences in mean scores (±S.D.)[a] of scales and single items of the QLQ-C30 and the QLQ-H&N35 between groups with different Karnofsky performance status (KPS) (data from first questionnaire only)

	KPS 40–80 ($n=184$)	KPS 90 ($n=181$)	KPS 100 ($n=257$)	P value[b]	RE
QLQ-C30					
Physical functioning	71±24.4	86±17.1	92±12.4	<0.001	31
Role functioning	65±36.1	83±25.0	91±19.2	<0.001	23
Emotional functioning	68±28.3	81±21.9	81±21.3	<0.001	9
Cognitive functioning	76±25.6	86±18.8	90±16.7	<0.001	12
Social functioning	76±28.8	85±22.9	90±17.9	<0.001	11
Fatigue	37±28.6	22±22.8	16±18.7	<0.001	23
Nausea/vomiting	9±17.7	5±11.6	2±8.2	<0.001	10
Pain	31±30.5	16±22.1	12±20.9	<0.001	19
Dyspnoea	32±34.6	19±25.6	12±22.5	<0.001	14
Insomnia	33±36.0	21±29.9	20±27.7	<0.001	6
Appetite loss	30±36.4	11±22.3	9±19.7	<0.001	16
Constipation	18±28.0	11±26.0	8±21.1	<0.001	8
Diarrhoea	7±18.9	5±16.3	3±12.2	0.01	3
Financial problems	20±31.2	12±24.8	10±23.1	<0.001	4
General QOL	56±22.0	71±21.2	75±21.9	<0.001	25
QLQ-H&N35					
Pain	25±25.0	17±20.0	13±19.3	<0.001	9
Swallowing	28±28.3	16±23.3	9±16.5	<0.001	22
Senses	27±32.5	14±24.6	10±20.7	<0.001	11
Speech	29±27.3	22±24.4	16±21.4	<0.001	8
Social eating	29±31.3	14±22.9	10±19.3	<0.001	22
Social contact	14±21.6	7±14.0	6±12.2	<0.001	9
Sexuality	37±39.1	28±35.1	20±28.3	<0.001	5
Teeth	24±35.9	17±27.0	16±27.1	0.2	1
Opening mouth	18±31.8	19±28.1	12±25.4	0.1	1
Dry mouth	43±39.2	35±35.0	37±35.8	0.2	1
Sticky saliva	39±38.0	31±34.4	27±33.6	0.006	3
Coughing	34±32.4	24±27.6	20±24.8	<0.001	7
Feeling ill	28±34.2	13±22.3	9±19.0	<0.001	13

RE, relative efficiency.
[a] A high score for a functional scale or global QOL implies a high level of functioning or global QOL, whereas a high score for a symptom score or single item implies a high level of symptoms.
[b] Kruskal–Wallis test.

Table 6
Means (±S.D.) of scores of scales and single items of the QLQ-H&N35 by symptom rating (data from first questionnaire only)

	Symptom rating[a]					
	0	1	2	3	4	P value[b]
QLQ-H&N35						
Pain	10±14.9	23±20.9	40±24.7	43±30.3	79±5.9	<0.001
	$n=364$	$n=148$	$n=56$	$n=26$	$n=2$	
Swallowing	8±12.9	27±21.4	29±25.5	68±23.6	77±13.4	<0.001
	$n=406$	$n=140$	$n=28$	$n=13$	$n=5$	
Senses	7±16.7	29±27.4	52±27.8	71±29.7		<0.001
	$n=444$	$n=84$	$n=39$	$n=30$		
Speech	14±19.7	24±1.7	41±26.8	53±28.7	44±31.4	<0.001
	$n=305$	$n=195$	$n=67$	$n=18$	$n=2$	
Dry mouth	15±24.2	43±29.0	70±27.2	89±18.6		<0.001
	$n=293$	$n=172$	$n=71$	$n=71$		

[a] For symptom rating, NCI toxicity criteria are used (pain, dysphagia, altered taste, speech and mouth dryness, respectively). $n<622$ due to missing data.
[b] One-way ANOVA.

Table 7
Change over time (before and after treatment) of mean scores (±S.D.)[a] of scales and single items of the QLQ-C30 and the QLQ-H&N35 (group I only, $n=232$)

	Before	After	P value[b]	SRM
QLQ-C30				
Physical functioning	84±20.6	74±24.6	<0.001	−0.54
Role functioning	77±31.9	62±34.2	<0.001	−0.45
Emotional functioning	71±25.7	72±24.8	0.7	0.03
Cognitive functioning	83±22.9	79±25.0	0.002	−0.22
Social functioning	82±25.5	76±28.6	<0.001	−0.25
Fatigue	28±25.9	43±28.6	<0.001	0.59
Nausea and vomiting	5±12.9	15±22.5	<0.001	0.44
Pain	24±27.8	33±30.3	<0.001	0.31
Dyspnoea	20±27.6	21±29.3	0.5	0.05
Insomnia	27±32.3	35±35.1	<0.001	0.23
Appetite loss	19±29.7	37±37.8	<0.001	0.43
Constipation	14±26.3	23±32.5	<0.001	0.31
Diarrhoea	6±15.2	8±19.7	0.06	0.13
Financial problems	12±25.2	18±30.4	<0.001	0.23
General QOL	62±23.6	54±23.1	<0.001	−0.32
QLQ-H&N35				
Pain	22±25.2	32±27.3	<0.001	0.28
Swallowing	17±25.1	37±28.6	<0.001	0.61
Senses	12±23.1	30±29.6	<0.001	0.68
Speech	26±26.7	40±29.1	<0.001	0.56
Social eating	16±24.2	34±27.9	<0.001	0.65
Social contact	9±15.9	18±24.1	<0.001	0.40
Sexuality	31±35.7	41±38.7	<0.001	0.27
Teeth	17±29.2	22±32.4	0.06	0.13
Opening mouth	16±29.5	32±36.1	<0.001	0.48
Dry mouth	28±31.1	47±38.3	<0.001	0.54
Sticky saliva	28±31.3	48±36.6	<0.001	0.63
Coughing	26±28.4	34±29.7	<0.001	0.23
Feeling ill	22±30.1	30±32.6	<0.001	0.27

SRM, standardised response mean.
[a] A high score for a functional scale or global QOL implies a high level of functioning or global QOL, whereas a high score for a symptom score or item implies a high level of symptoms.
[b] Mann–Whitney test.

The questionnaire was well accepted by the patients and the compliance was high. The number of missing items was generally very low. As in other modules [4], the sexuality items were problematic. Nevertheless, approximately 90% of patients answered both questions and few indicated that these questions were inappropriate or upsetting. The main reason that patients did not answer these questions was because they were sexually inactive. The time needed for completing both questionnaires (<20 min in 95% of patients) is very acceptable and makes it feasible to use them in clinical studies. Approximately one-quarter (26%) of the patients needed help, often consisting of help to read the questions because the patients did not have reading glasses. In our opinion, this does not mean that the QLQ-H&N35 is inappropriate or unacceptable. Patients who were unable to understand the questionnaire (because they were senile/demented, had severe cognitive impairment, or were illiterate) were excluded from the study. Measurement of HQOL by proxy might be considered for these patients [28], but this has not yet been approved as an acceptable alternative.

Scaling analysis showed the scale structure to perform very well, indicating a high level of construct validity. Few scaling errors were observed. The main problem was the item about painful throat. From a clinical perspective, it is understandable that this item had a higher correlation with the swallowing scale. However, since this item represents a specific type of pain which is relevant for H&N cancer patients, in particular those with pharyngeal cancer, it was decided for clinical reasons to leave this item in the pain scale. The QLQ-H&N35 and the QLQ-C30 both contain a pain scale and a social functioning scale; the correlation between the corresponding scales was moderately high. In addition, taking into account the wording of the items, the pain and social contacts scale of the QLQ-H&N35 seem to add a specific H&N dimension to the use of the pain and social functioning scale of the QLQ-C30.

According to the preliminary analysis [10] some items (problems with teeth and the last five items) were candidates for removal. With regard to the teeth item, from a clinical point of view this is regarded as an important item in patients with oral and oropharyngeal cancer.

There were clear differences between sites. Therefore, for reasons of content validity, it was decided to keep this item in the questionnaire. With regard to the last five items, those can be considered optional. They can be omitted in studies, in which these data can be reliably collected by other means. However, it should be realised that clinicians may not always be aware of the use of pain killers and nutritional supplements (in particular those which can be bought without prescription) and therefore, we have still included them in the questionnaire.

The reliability as assessed by Cronbach's alpha coefficient (internal consistency) was excellent. A test–retest procedure was not performed in this study. Test–retest data are available from an analysis of 120 Swedish patients 3 years after primary treatment of head and neck cancer (E. Hammerlid, Sahlgrenska Hospital, Goteborg, data not shown). In that study, the intraclass correlations ranged from 0.76 (senses) to 0.94 (social eating) for the scales and from 0.65 (feeling ill) to 0.86 (dry mouth) for the single items of the QLQ-H&N35. There was only one *r*-value below 0.70 (feeling ill). Thus, the test–retest reliability seems to be comparable with that of the QLQ-C30 [29].

There were very clear-cut differences in the scores of almost all scales and single items of the QLQ-H&N35 between disease states, site and patients with different Karnofsky performance status. The differences between different disease states clearly reflect differences between tumour-related symptoms (patients with newly diagnosed disease), long-term complications of treatment (disease-free patients) and the combination of recurrence- and treatment-related symptoms (patients with recurrent disease). For example, pain was most prominent in patients with newly diagnosed or recurrent disease, whereas dry mouth and sticky saliva (a long-term complication of radiotherapy to the region of the salivary glands) were most prominent in disease-free patients and patients with recurrent disease. Differences between sites also reflected site-specific differences in symptoms (e.g. high level of speech problems in laryngeal cancer; high level of swallowing problems in oral and pharyngeal cancer; only small differences in social contact and sexuality problems between sites). Both the QLQ-C30 and the QLQ-H&N35 showed a strong correlation between the scores of scales or single items and the Karnofsky performance status. The only area where the QLQ-H&N35 failed to show a consistent correlation pattern was in disease stage. The expected correlation between more advanced stages and higher levels of symptoms in patients with newly diagnosed disease was not found for the group as a whole or for specific sites. This is contrary to the results of the previous analysis [10]. The lack of this correlation may be due to the low level of the symptom scores of the QLQ-H&N35 in patients with newly diagnosed disease in this sample, even in advanced stages, and to the fact that patients requiring a primary laryngectomy (reflecting very advanced disease) were not included in this study.

The QLQ-H&N35 was able to detect significant deterioration of symptoms after treatment. The differences were in the order of 10–20; for the QLQ-C30 this has been shown to indicate a clinically significant effect [23].

The QLQ-C30 was also able to detect significant differences between disease status, sites and patients with different Karnofsky performance status, although for disease status and site the differences were less pronounced than those found with the QLQ-H&N35. This demonstrates the validity of the QLQ-C30 in H&N cancer patients, as was shown before, [11] but it also illustrates the need for a H&N cancer-specific module to increase the possibility of detecting differences in H&N cancer-specific HQOL [30]. The specificity of the QLQ-H&N35 was also illustrated by its sensitivity as measured by the relative efficiency. When differences between disease status and sites were studied, higher relative efficiency values were seen for the physical symptom scales and items of the QLQ-H&N35 compared with the scales and items of the QLQ-C30.

Version 3.0 of the QLQ-C30 differs from the previous version with regard to the response format of the physical functioning scale. It was to be expected that the scale with the four-point scale would have better reliability than the version 1.0 scale, and the values of Cronbach's alpha coefficient and the item–scale correlations confirm this. Similarly, we confirmed that the new scale was more strongly associated with Karnofsky performance status scores. The means and standard deviations for both physical functioning scales happen to be similar in both studies, but this may well be due to chance in these two studies; the new scale was not designed with this specifically in mind. Thus, the anticipated improvements in the new scale were confirmed, and we conclude that the QLQ-C30 version 3.0 should be used as the current standard in all future studies.

The QLQ-H&N35 has now proven its value in the assessment of HQOL in H&N cancer patients having been tested more extensively than other instruments. The Functional Assessment of Cancer Therapy-H&N scale (FACT-HN) [31] is used in the same way (in conjunction with a general cancer module) and has been translated in various languages. However, the data on reliability and validity of the FACT-HN are limited [31,32]. Other instruments have been tested less extensively [33–36] and/or are aimed at specific subgroups, e.g. patients receiving radiotherapy [37,38]. Moreover, the QLQ-H&N35 has been tested in 12 countries and nine languages.

Thus, the EORTC QLQ-H&N35, in conjunction with the QLQ-C30, can be regarded as a standard instrument to measure HQOL in H&N cancer patients. Copies of the questionnaire and scoring instructions can be obtained from the Quality of Life Unit of the EORTC Data Center in Brussels.

Acknowledgements

The following investigators also recruited patients into this study: C. Baughan (General Hospital, Southamptom, UK), A. Cull (Western General Hospital, Edinburgh, UK), L. Feenstra (Universiteitsziekenhuis Saint Rafael, Leuven, Belgium), K. Gamble and A. Bartlet (Royal Marsden Hospital, London, UK) and K. Pekkola-Heino (University Central Hospital, Turku, Finland). We are grateful for their contribution.

Appendix. EORTC QLQ-H&N35

Patients sometimes report that they have the following symptoms or problems. Please indicate the extent to which you have experienced these symptoms or problems during the past week. Please answer by circling the number that best applies to you.

		Not at all	A little	Quite a bit	Very much
During the past week:					
1.	Have you had pain in your mouth?	1	2	3	4
2.	Have you had pain in your jaw?	1	2	3	4
3.	Have you had soreness in your mouth?	1	2	3	4
4.	Have you had a painful throat?	1	2	3	4
5.	Have you had problems swallowing liquids?	1	2	3	4
6.	Have you had problems swallowing pureed food?	1	2	3	4
7.	Have you had problems swallowing solid food?	1	2	3	4
8.	Have you choked when swallowing?	1	2	3	4
9.	Have you had problems with your teeth?	1	2	3	4
10.	Have you had problems opening your mouth wide?	1	2	3	4
11.	Have you had a dry mouth?	1	2	3	4
12.	Have you had sticky saliva?	1	2	3	4
13.	Have you had problems with your sense of smell?	1	2	3	4
14.	Gave you had problems with your sense of taste?	1	2	3	4
15.	Have you coughed?	1	2	3	4
16	Have you been hoarse?	1	2	3	4
17.	Have you felt ill?	1	2	3	4
18.	Has your appearance bothered you?	1	2	3	4
19.	Have you had trouble eating?	1	2	3	4
During the past week:					
20.	Have you had trouble eating in front of your family?	1	2	3	4
21.	Have you had trouble eating in front of other people?	1	2	3	4
22.	Have you had trouble enjoying your meals?	1	2	3	4
23.	Have you had trouble talking to other people?	1	2	3	4
24.	Have you had trouble talking on the telephone?	1	2	3	4
25.	Have you had trouble having social contact with your family?	1	2	3	4
26.	Have you had trouble having social contact with friends?	1	2	3	4
27.	Have you had trouble going out in public?	1	2	3	4
28.	Have you had trouble having physical contact with family or friends?	1	2	3	4
29.	Have you felt less interest in sex?	1	2	3	4
30.	Have you felt less sexual enjoyment?	1	2	3	4
During the past week:				No	Yes
31.	Have you used pain killers?			1	2
32.	Have you taken any nutritional supplement (excluding vitamins)?			1	2
33.	Have you used a feeding tube?			1	2
34.	Have you lost weight?			1	2
35.	Have you gained weight?			1	2

Scales and single items

Pain	Items 1, 2, 3, 4
Swallowing	Items 5, 6, 7, 8
Senses	Items 13, 14
Speech	Items 16, 23, 24
Social eating	Items 19, 20, 21, 22
Social contact	Items 18, 25, 26, 27, 28
Sexuality	Items 29, 30
Single items	Items 9, 10, 11, 12, 15, 17, 31, 32, 33, 34, 35

References

1. Morris J. Widening perspectives: quality of life as a measure of outcome in the treatment of patients with cancers of the head and neck. *Oral Oncol* 1994, **30B**, 29–31.
2. The Department of Veteran Affairs Laryngeal Cancer Study Group. Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. *N Engl J Med* 1991, **324**, 1685–1690.
3. Aaronson NK, Cull A, Kaasa S, *et al*. The European Organization of Research and Treatment of Cancer (EORTC) modular approach to quality of life assessment in oncology: and update. In Spilker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials* 2nd edn. New York, NY, Raven Press, 1996, 179–189.
4. Sprangers MAG, Groenvold M, Arraras JI, *et al*. The European Organization for Research and Treatment of Cancer breast cancer-specific module: first results from a three-country field study. *J Clin Oncol* 1997, **14**, 2756–2768.
5. Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. 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, **30A**, 635–642.
6. 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.
7. Sprangers MAG, Cull A, Bjordal K, *et al*. 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.
8. Bjordal K, Ahlner-Elmqvist M, Tolleson 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. *Acta Oncol* 1994, **33**, 879–885.
9. Cull A, Sprangers MAG, Bjordal K, *et al*. *EORTC Quality of Life Study Group Translation Procedure*. Brussels, Belgium, European Organization for Research and Treatment of Cancer, 1998.
10. Bjordal K, Hammerlid E, Ahlner-Elmqvist M, *et al*. Quality of life in head and neck cancer patients: validation of the EORTC QLQ-H&N35. *J Clin Oncol* 1999, **17**, 1008–1019.
11. 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.
12. Bjordal K, Kaasa S. Psychological distress in head and neck cancer patients 7–11 years after curative treatment. *Br J Cancer* 1995, **71**, 592–597.
13. Hammerlid E, Mercke C, Sullivan M, Westin T. A prospective quality of life study of patients with oral or pharyngeal carcinoma treated with external beam irradiation with or without brachytherapy. *Oral Oncol* 1997, **33**, 189–196.
14. Hammerlid E, Bjordal K, Ahlner-Elmqvist M, *et al*. Prospective, longitudinal quality-of-life study of patients with head and neck cancer: a feasibility study including the EORTC QLQ-C30. *Otolaryngol Head Neck Surg* 1997, **116**, 666–673.
15. Hammerlid E, Mercke C, Sullivan M, Westin T. A prospective quality of life study of patients with laryngeal carcinoma by tumor stage and different radiation therapy schedules. *Laryngoscope* 1998, **108**, 747–759.
16. De Graeff A, de Leeuw JRJ, Ros WJG, Hordijk GJ, Blijham GH, Winnubst JAM. A prospective study on quality of life of patients with cancer of the oral cavity or oropharynx treated with surgery with or without radiotherapy. *Oral Oncol* 1999, **35**, 27–32.
17. De Graeff A, de Leeuw JRJ, Ros WJG, *et al*. A prospective study on quality of life of patients with laryngeal cancer treated with radiotherapy. *Head Neck* 1999, **21**, 291–296.
18. Karnofsky DA, Burchenal JH. The evaluation of chemotherapeutics in cancer. In Macleod CM, ed. *Evaluation of Chemotherapeutic Agents*. New York, Columbia, University Press, 1949, 191–205.
19. National Cancer Institute. *NCI Common Toxicity Criteria*. New York, National Cancer Institute, 1998.
20. Nunnally JC, Bernstein IH. *Psychometric Theory*. New York, McGraw-Hill, 1994.
21. Hays RD, Hayashi T, Carson S, *et al*. *User's Guide for the Multitrait Analysis Program (MAP)*. Santa Monica, CA, The RAND Cooperation, 1988.
22. Kerlinger FN. *Foundations of Behavioural Research*. New York, Holt, Rinehart and Winston, 1973.
23. 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.
24. Norusis MJ. *SPSS for Windows*. Chicago, SPSS Inc, 1993.
25. Fayers P, Machin D. *Quality of Life: Assessment, Analysis and Interpretation*. Chichester, John Wiley, 2000, 66–71.
26. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ, Lawrence Erlbaum, 1988.
27. StataCorp. *Stata Statistical Software: Release 5.0*. College Station, TX, 1997.
28. Sneeuw KCA, Aaronson NK, Sprangers MAG, *et al*. *J Clin Oncol* 1997, **15**, 1206–1217.
29. Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S. Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-life Questionnaire. *J Clin Oncol* 1998, **16**, 1188–1196.
30. Gliklich RE, Goldsmith TA, Funk GF. Are head and neck specific quality of life measures necessary? *Head Neck* 1997, **19**, 474–480.
31. List MA, D'Antonio LL, Cella DF, *et al*. The performance status for head and neck cancer patients and the functional assessment of cancer therapy-head and neck scale: a study of utility and validity. *Cancer* 1996, **77**, 2294–2301.
32. D'Antonio LL, Zimmerman GJ, Cella DF, Long SA. Quality of life and functional status measures in patients with head and neck cancer. *Arch Otolaryngol Head Neck Surg* 1996, **122**, 482–487.
33. Rathmell AJ, Ash DV, Howes M, *et al*. Assessing quality of life in patients treated for advanced head and neck cancer. *Clin Oncol* 1991, **3**, 10–16.
34. Hassan SJ, Weymuller EA. Assessment of quality of life in head and neck cancer patients. *Head Neck* 1993, **15**, 485–496.
35. Baker C. A functional status scale for measuring quality of life outcomes in head and neck cancer patients. *Cancer Nursing* 1995, **18**, 452–457.
36. Terrel JE, Nanavati KA, Esclamado RM, Bishop JK, Bradford C, Wolf GT. Head and neck-specific quality of life. Instrument validation. *Arch Otolaryngol Head Neck Surg* 1997, **123**, 1125–1132.
37. Browman GP, Levine MN, Hodson DI, *et al*. The head and neck radiotherapy questionnaire: a morbidity/quality-of-life instrument for clinical trials of radiation therapy in locally advanced head and neck cancer. *J Clin Oncol* 1993, **11**, 863–872.
38. Trotti A, Johnson DJ, Gwede C, *et al*. Development of a head and neck companion mode for the quality of life-radiation therapy instrument (QOL-RTI). *Int J Radiat Oncol Biol Phys* 1998, **42**, 257–261.