Pergamon

European Journal of Cancer, Vol. 35, No. 2, pp. 238–247, 1999

Printed in Great Britain
0959-8049/99/$ - see front matter

PII: S0959-8049(98)00357-8

Original Paper

The Construction and Testing of the EORTC Colorectal Cancer-specific Quality of Life Questionnaire Module (QLQ-CR38)

M.A.G. Sprangers,[1] A. te Velde,[2] and N.K. Aaronson[2,3] on behalf of the European Organization for Research and Treatment of Cancer Study Group on Quality of Life

[1]Department of Medical Psychology, J404, Academic Medical Centre, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam; [2]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam; and [3]Department of Medical Psychology, Free University, Amsterdam, The Netherlands

The objectives of the current study were to construct a colorectal cancer-specific quality of life (QL) questionnaire module to be used in conjunction with the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 and to test its reliability and validity in The Netherlands. Module construction took place following the EORTC guidelines for module development. The module—the QLQ-CR38—consists of 38 items covering symptoms and side-effects related to different treatment modalities, body image, sexuality and future perspective. This module was tested among 117 colorectal cancer patients on several occasions. The timing was prior to treatment with radiotherapy or chemotherapy, during treatment and 3 months following the second assessment. For purposes of test–retest reliability, a subsample of patients completed the QLQ-CR38 1 week following the third assessment. Multitrait scaling analysis confirmed the hypothesised scale structure of the function scales but not of the symptom scales. Cronbach's alpha coefficients for seven of the nine scales exceeded the 0.70 criterion at one or both assessments. The test–retest reliability for all scales and one single item was 0.78 or higher. The stability of the two remaining single items was lower. On the basis of known-groups comparisons, selective scales distinguished clearly between patients differing in disease stage, initial and on-treatment performance status and the presence of a stoma. Additionally, selective scales detected change over time as a function of change in performance status and treatment-induced change. These results lend support to the clinical validity of the QLQ-CR38 as a supplementary questionnaire for assessing specific QL issues relevant to patients with colorectal cancer. Additional efforts to test the module's cross-cultural validity are needed.

Key words: colorectal neoplasms, quality of life, assessment, EORTC, validation
Eur J Cancer, Vol. 35, No. 2, pp. 238–247, 1999

INTRODUCTION

COLORECTAL CANCER is a major health problem in Western societies. It is currently the second and third most common cancer among women and men, respectively, in the majority of the countries of western Europe and in the U.S.A. [1]. A literature review of the quality of life (QL) of colorectal cancer patients undergoing surgery indicated that none of the studies employed standardised QL questionnaires that were tailored to the needs of colorectal cancer patients [2]. This review was based on studies published up to 1992, but a more recent review revealed that up to 1997, standardised colorectal cancer-specific questionnaires were still not used [3]. We therefore developed a QL instrument specifically for use among colorectal cancer patients. This work was carried out in the context of a long-term project of the European Organization for Research and Treatment of Cancer (EORTC) Study Group on Quality of Life to develop an

Correspondence to M.A.G. Sprangers, e-mail: m.a.sprangers@amc.uva.nl
Received 26 Jan. 1998; revised 19 Aug. 1998; accepted 8 Oct. 1998.

integrated measurement system for evaluating the QL of cancer patients participating in international clinical trials. The Study Group has adopted a modular measurement approach to QL assessment whereby a 'core' instrument (the QLQ-C30) [4] has been designed to cover a range of QL issues relevant to a broad spectrum of cancer patients. The QLQ-C30 is intended to be supplemented by more specific questionnaire modules to assess aspects of QL of particular importance to various patient subgroups. Detailed guidelines have been established for generating diagnosis- and/or treatment-specific, supplemental questionnaires to be used in conjunction with the core instrument [5, 6]. Modules related to brain cancer [7], breast cancer [8], head and neck cancer [9], lung cancer [10], oesophageal cancer [11] and ovarian cancer [12] have been constructed following these guidelines.

The primary objectives of the current study were to construct a colorectal cancer-specific questionnaire module to be used in conjunction with the EORTC QLQ-C30 for assessing the QL of patients participating in international clinical trials and to test the psychometric performance (i.e. reliability, validity and responsiveness) of this module among Dutch colorectal cancer patients.

THE CONSTRUCTION OF THE EORTC COLORECTAL CANCER MODULE

The colorectal cancer module is intended for use among a wide range of colorectal cancer patients participating in international clinical trials. The module was therefore meant to include primarily disease symptoms and side-effects related to different treatment modalities (i.e. sphincter saving resection, rectum extirpation, radiotherapy and chemotherapy). Additionally, QL aspects that can be affected by (surgical) treatments (i.e. body image and sexuality) or that are related to stage of disease (i.e. future perspective) were also considered for inclusion. The development of this module was a collective effort of primarily Dutch and American members of the Study Group.

In order to compile an exhaustive list of relevant QL issues that cover the identified domains, a series of literature searches was conducted on the MEDLINE database for the years 1969–1992. A list of 65 issues was thus generated and presented to five cancer specialists (i.e. two gastroenterologists, one surgeon, one radiotherapist and one stoma nurse) from the Antoni van Leeuwenhoek Hospital, a cancer treatment centre in Amsterdam, The Netherlands. On the basis of their comments regarding the appropriateness of content and the breadth of coverage, a revised list containing 52 issues was devised.

Questions were then developed based on these issues according to a range of criteria. For example, questions should preferably be compatible with the response categories adopted for the QLQ-C30 (i.e. from (1) 'not at all' to (4) 'very much'), be compatible with respect to the 1 week time-frame and refer to the presence of a complaint rather than to the trouble it may have caused [13]. The items on body image were selected from a 10 item scale devised by Hopwood on the basis of their content validity [14]. These items were also incorporated in the EORTC breast cancer module [8]. The resulting, provisional questionnaire module contained 60 items. The wording of these items was discussed during plenary sessions of the Study Group. The involvement of individuals representing a broad range of cultures and languages helped to ensure that the wording of the questionnaire items remained relatively simple, straightforward and free of jargon. Whilst the discussions were held in English, the first version of the module was written in Dutch.

Patient interviews

The first version of the questionnaire was pretested among Dutch patients to identify and solve potential problems in its administration. However, since patients were not consulted during the previous QL issue generation phase, the pretesting interviews were expanded to elicit patients' opinions about the relevance of the topics covered by the provisional module. The interviews started with an open question asking about patients' experiences with their health. Then the QLQ-C30 and the provisional module were administered and patients were asked to report the extent to which they had experienced the problems enumerated. Patients were subsequently asked to indicate any significant omissions in the list, to comment on the wording of each item and to choose a preset number of items which, in their opinion, should definitely be included in the final questionnaire. These interviews were conducted with each patient individually. The sample included 24 colorectal cancer patients undergoing either radiotherapy subsequent to surgery for local/locoregional disease or chemotherapy for metastatic disease. 11 patients had a permanent stoma. All patients were treated at the Antoni van Leeuwenhoek Hospital.

The module was well received as patients found the questionnaire relevant and easy to answer. On the basis of qualitative and quantitative findings, the questionnaire was revised, according to preset criteria. For example, items were deleted if they had a low mean score (< 1.5), a small range of responses (< 2 points), a low prevalence (< 0.30) and a small number of priority ratings (i.e. less than one-third of the patients indicating that the item should definitely be retained in the final questionnaire). Additionally, on the basis of patients' comments, a number of items were (slightly) reworded and one item relating to belching was added [15].

Resulting module and hypothesised scale structure

The resulting module—the QLQ-CR38—consists of 38 items. Nineteen questions are completed by all patients whilst the remaining 19 questions are completed by subsamples of patients (i.e. males or females, patients with or without a stoma). The content of the module, organised by its hypothesised scale structure, is provided in the Appendix. The QLQ-CR38 incorporates two functional scales (body image and sexuality) and seven symptom scales (micturition problems, symptoms in the area of the gastrointestinal tract, chemotherapy side-effects, problems with defaecation, stoma-related problems, male and female sexual problems). Because the item on 'sexual enjoyment' is conditional on having been sexually active, it is completed by only a subset of patients, and is therefore handled singly. The remaining single items assess future perspective and weight loss. All items employ the four-category response option that is used in the QLQ-C30. The items have a 1 week time-frame, with the exception of the items on sexuality. Since patients may engage in sexual activities less frequently than once a week, a 4 week time-frame was chosen. This version of the module was approved by the Study Group and was used in the current study.

The following sections report the results of testing the reliability and validity of the QLQ-CR38 when used among Dutch patients with colorectal cancer.

PATIENTS AND METHODS

Patients

Colorectal cancer patients with local/locoregional disease receiving radiotherapy and patients with inoperable, locally advanced disease or with distant metastases receiving chemotherapy (i.e. ICI D1694 or 5-fluorouracil (5-FU) with or without leucovorin or levamisol) were recruited from the Antoni van Leeuwenhoek Hospital in Amsterdam. Patients were excluded who: (a) had a life expectancy < 3 months; (b) had had a liver resection; (c) lacked basic proficiency in Dutch; or (d) were participating in concurrent QL investigations. No restrictions were made with regard to age or performance status.

Additional study measures

The QLQ-CR38 was administered together with the EORTC QLQ-C30 [4], the CARES-SF [16], and the SF-36 [17]. The data related to these other questionnaires are reported elsewhere [18, 19]. The order in which these questionnaires were administered to patients was varied randomly to avoid any systematic order effects. Sociodemographic data included age, gender, marital status, education and employment status. Clinical data included disease stage, nature and schedule of treatment and the presence of a stoma. This information was extracted from the patients' medical records. To establish the level of patients' performance status, a short interview was conducted with the patients based, in part, on guidelines recommended by Schag and colleagues [20]. The response patterns to the 11 interview items were rescored to obtain a Karnofsky performance status score [21]. Finally, a 'debriefing' form was used to ascertain the time needed for completion of the questionnaire, the degree and kind of help provided, the presence of items that were confusing, difficult to understand, or intrusive, and whether patients had skipped questions and, if so, the reason(s) why.

Data collection procedure

The QLQ-C30, the colorectal cancer module, and the debriefing questions were administered in this order on several occasions. Patients receiving radiotherapy completed these measures prior to their treatment and on the last day of the first course of radiotherapy. Patients receiving chemotherapy completed the questionnaires on the first day of the first treatment cycle (prior to receiving the chemotherapy) and on the first day of one of the subsequent cycles (i.e. from the second to the seventh cycle), dependent on the treatment schedule. The questionnaires were self-administered either at the outpatient clinic or on the hospital ward. A research assistant was present during the assessments to provide assistance, if needed, and to record any problems patients may have had with completing the questionnaire.

A third assessment was carried out 3 months after the second assessment. At this assessment, the sample was randomly divided into three different mode of administration conditions: one-third of the patients, again, completed the questionnaire during their scheduled visit to the hospital whilst a research assistant was present, one-third answered all questions in a telephone interview and one-third completed the questionnaire by mail. To establish test–retest reliability, a random subsample of patients completed the QLQ-CR38 1 week following the third assessment. For practical reasons the test–retest sample was limited to those patients in the mail and telephone conditions. The mode of administration at the third and fourth assessments was held constant. Data of these third and fourth assessments will be presented in so far as they pertain to test–retest reliability, whilst mode of administration will not be further addressed in this paper.

Statistical analyses

A range of analyses was carried out to test empirically the hypothesised scale structure of the QLQ-CR38, to establish scale reliability and to evaluate the validity of the questionnaire scales and single items. These analyses were conducted on the data of the first and second assessments.

Multitrait scaling. Multitrait scaling analysis was employed to examine the extent to which the items of the questionnaire could be combined into the hypothesised multi-item scales. The technique is based on an examination of item–scale correlations [22]. Evidence of item internal consistency is defined as a correlation of ≥ 0.40 (corrected for overlap) between an item and its own scale. Item discriminant validity is indicated when an item correlates significantly higher with its own scale (corrected for overlap) than with another scale (referred to as scaling success).

Reliability. The internal consistency of the multi-item questionnaire scale was assessed by Cronbach's alpha coefficient [23]. Internal consistency estimates of a magnitude of 0.70 were sought [24].

The stability or test–retest reliability data are based on two modes of administration (mail and telephone at the third and fourth assessments) that differ from the in-clinic procedure at the first two assessments. We therefore first examined the effect of administration procedures on the data of the third assessment. Mean scale score differences were tested for statistical significance by means of one-way analyses of covariance, with age, stage and treatment used as covariates. Since the alternative and the null hypothesis concur (i.e. no differences among administration procedures), the P value was set at 0.10 to employ a conservative criterion. The stability of the multi-item scales and single items was assessed by means of the intraclass correlation coefficient (ICC) [25].

Validity. The validity of the QLQ-CR38 was assessed in two ways. First, the method of known-groups comparison [26] was employed to assess the ability of the QLQ-CR38 to distinguish between subgroups of patients differing in clinical status. The clinical parameters employed to form mutually exclusive patient subgroups for the baseline analyses included disease stage (local/locoregional versus metastatic disease) and initial performance status (Karnofsky performance status $\geq$ 80 versus Karnofsky performance status ≤ 70). For the analysis of the data from the second administration, the clinical parameters used were on-treatment performance status and the presence of a stoma (stoma versus non-stoma). One-way analysis of variance (ANOVA) was used to test for the statistical significance of group differences. To examine the magnitude of such differences, effect sizes based on standardised differences between mean scores were calculated. Following Cohen [27], effect sizes of 0.20, 0.50 and 0.80 were considered small, medium and large, respectively.

Second, the responsiveness of the QLQ-CR38 to observed changes in performance status (Karnofsky performance status scores) was assessed. Improvement or deterioration in performance status was defined as a shift of at least one level upward or downward on the Karnofsky performance status. Additionally, changes in the QLQ-CR38 scores over time were examined in relationship to treatment-induced change.

We expected patients who had received radiotherapy to report higher levels of radiotherapy-related symptoms (i.e. micturition problems, symptoms in the area of the gastrointestinal tract and problems with defaecation) at the second assessment than at baseline. Conversely, patients undergoing chemotherapy were expected to report higher levels of chemotherapy-related side-effects. Repeated-measures ANOVA was used to test for statistically significant changes in QLQ-CR38 scores over time. Only statistically significant ($P < 0.05$) results are reported.

RESULTS

Patient recruitment and follow-up

From September 1992 to April 1994, 145 patients meeting the eligibility criteria were invited to participate in the study. 28 patients (19%) declined because the study was perceived as too confronting or too burdensome ($n = 14$), a perceived lack of time ($n = 4$), or the patient felt too ill ($n = 2$). The other 8 patients had other reasons or their reasons for refusal were not recorded. Patients declining participation were, on average, older (mean age 68 years versus 61 years; $P < 0.01$), were less frequently married (61% versus 81%; $P = 0.06$) and more often had only received compulsory education (66% versus 58%; $P = 0.1$), than those who participated.

Follow-up. At the second assessment point, 8 patients were lost to follow-up for various reasons, including feeling too ill and perceived burden. The average time between the first and second assessments was 32.1 days (standard deviation, S.D. = 10.5 days). At the third assessment point, 14 patients were lost to follow-up, primarily because of severe illness or death ($n = 10$). A random subsample of 26 patients was asked to complete the questionnaire a fourth time. 1 patient declined and another patient was missed because of administrative failure. The average time between the third and fourth assessments was 6.5 days (S.D. = 3.9 days).

Sociodemographic and clinical data

The sociodemographic and clinical characteristics of the 117 patients at baseline are presented in Table 1. The patients varied in age between 35 and 86 years, with a mean of 61 years (S.D. = 11 years). The majority of the patients were male, were married, had compulsory education only, were unemployed (including retirement), received radiotherapy and had intact sphincters. The sample was heterogeneous with regard to extent of disease and pretreatment Karnofsky performance status.

Feasibility of the questionnaire

The average time required to complete the QLQ-CR38 in combination with the QLQ-C30 at baseline was 10.3 min (S.D. = 5.6 min). 95/115 (83%) patients were able to complete the questionnaire without assistance (for 2 patients these feasibility data were missing). The questionnaire was administered in interview form in 8 cases, while help was provided in 12 cases (e.g. the interviewer circled the answers for the patient). These latter patients were older, had a worse Karnofsky performance status score and a lower level of education (data not shown).

In general, the questions were well accepted and clear to the majority of patients. The only items that elicited a number of comments were those related to sexuality. Between 3% of the men and 12% of the women declined to complete these items at baseline, primarily because they were considered too intrusive.

Scaling and reliability estimation

Multitrait scaling. A summary of the results of the multitrait scaling analyses across the two assessment points is provided in Table 2. All item–rest correlations exceeded the 0.40 criterion for item internal consistency for the body image scale and the three sexuality scales at the two assessments, and for the micturition problems scale at follow-up. Although the remaining scales displayed some problems, the chemotherapy side-effects scale was the only scale whose item–rest correlations were all smaller than 0.40 at baseline. The mean item–rest correlation across all nine scales was 0.55 at baseline, and 0.62 at the second administration of the questionnaire.

For each assessment there were 272 (34 items $\times (9 - 1)$ scales) tests of item-discriminant validity. Scaling successes were higher for the sexual functioning scale (100%), the body image scale (range 92–100%), the male sexual problems scale (range 50–100%) and the micturition problems scale (range 67–83%) and lower for the defaecation problems scale (range 57–71%), female sexual problems scale (range 50–58%), chemotherapy side-effects scale (0–67%), gastrointestinal tract symptoms scale (30–40%) and stoma-related problems scale (range 25–39%). For the nine scales combined, scaling successes were noted in 58% of the cases at the first assessment and in 68% of the cases at follow-up.

Table 1. *Sociodemographic and clinical characteristics of the study sample at baseline*

Characteristics	$n = 117$
Age (years)	
Mean (S.D.)	61 (11)
Range	35–86
Gender	
Male	74 (63%)
Female	43 (37%)
Marital status	
Unmarried	8 (7%)
Married	94 (80%)
Divorced	2 (2%)
Widowed	12 (10%)
Unknown	1 (1%)
Education	
Compulsory only	67 (57%)
Advanced vocational	24 (21%)
University	25 (21%)
Unknown	1 (1%)
Employed	
Yes	32 (27%)
No*	81 (69%)
Unknown	4 (3%)
Disease stage	
Local	24 (21%)
Loco-regional	51 (44%)
Metastatic	42 (36%)
Treatment	
Chemotherapy	44 (38%)
Radiotherapy	73 (62%)
Stoma	
Yes	46 (39%)
No	71 (61%)
KPS score	
Mean (S.D.)	79.3 (12.6)
Range	40–100

*Includes retirement. S.D., standard deviation; KPS, Karnofsky performance status.

Table 2. Item internal consistency (IN), item internal consistency test (INT), item discriminant validity (DIS) and item discriminant validity test (DIST) at baseline (T1) and follow-up (T2)

Scales	Items	*n*	IN*	INT†	DIS‡	DIST§
BI T1	3	74	0.47–0.63	100%	0.00–0.34	92%
BI T2		72	0.62–0.71	100%	0.06–0.43	100%
SX T1	2	64	0.78	100%	0.03–0.31	100%
SX T2		64	0.73	100%	0.02–0.31	100%
MI T1	3	74	0.29–0.64	67%	0.05–0.43	83%
MI T2		72	0.46–0.63	100%	0.01–0.44	67%
GI T1	5	74	0.13–0.63	60%	0.01–0.47	40%
GI T2		72	0.20–0.41	20%	0.06–0.47	30%
CT T1	3	74	0.16–0.26	0%	0.01–0.35	0%
CT T2		72	0.38–0.48	67%	0.00–0.38	67%
DF T1	7	74	0.01–0.67	71%	0.03–0.57	57%
DF T2		72	0.31–0.69	71%	0.02–0.46	71%
STO T1	7	42	0.30–0.76	57%	0.04–0.46	39%
STO T2		36	0.28–0.76	71%	0.00–0.56	25%
MSX T1	2	43	0.63	100%	0.04–0.53	50%
MSX T2		44	0.86	100%	0.02–0.27	100%
FSX T1	2	8	0.91	100%	0.00–0.87	58%
FSX T2		9	0.94	100%	0.00–0.71	50%

Since a number of questions were completed by a subsample of patients (males or females, patients with or without a stoma) a series of multitrait scaling analyses was conducted including different combinations of scales. The results are reported from those five series of analyses that contained the largest number of patients. BI, body image; SX, sexual function; MI, micturition problems; GI, gastrointestinal tract symptoms; CT, chemotherapy side-effects; DF, defaecation problems; STO, stoma-related problems; MSX, male sexual problems; FSX, female sexual problems. *In this column the range of item–rest correlations is presented. †Item internal consistency test = percentage of item–scale correlations (corrected for overlap) greater than 0.40. ‡In this column the range of correlations between items and all other scales is presented. §Item discriminant validity test = percentage of cases where item–rest correlations are significantly higher than the correlations between the items and the other scales.

Internal consistency reliability. Cronbach's coefficient alpha for the multi-item scales exceeded the 0.70 criterion for the body image, sexual functioning, defaecation problems, stoma-related problems and male sexual problems scales at both administrations, and for micturition problems and female sexual problems scales at the second assessment. The coefficients for the chemotherapy side-effects scale (range 0.47–0.63) and the gastrointestinal tract symptoms scale (range 0.49–0.61) were lower (Table 3). No systematic differences in scale reliabilities were found as a function of patients' age, educational level, performance status, or stage of disease (data not shown).

Test–retest reliability. The mode of administration only had a statistically significant ($P < 0.10$) impact on the reported level of sexual functioning and male sexual problems. In these cases there was no systematic pattern of differences in the mean scores across the three administration procedures.

The majority of the QLQ-CR38 scales exhibited good to excellent stability, with ICCs ranging from 0.78 (gastrointestinal tract symptoms) to 0.92 (sexual functioning) (see Table 3). The stability of two single items (future perspective and weight loss) was lower (0.53 and 0.55, respectively). The ICCs of the mail and telephone condition separately, were comparable despite the small subsample sizes (data not shown).

Score distributions

Following the scoring procedures for the QLQ-C30, all scale and single-item scores were linearly transformed to a 0–100 scale. For the functional scales and single items (i.e. body image, sexual functioning, sexual enjoyment and future perspective), higher scores represent a higher level of functioning. For the symptom scales and single items, a higher score represents a higher level of symptoms.

Table 3 depicts the means and S.D.s for the baseline and second assessments. For the scales or individual items assessing body image, sexual functioning, sexual enjoyment and future perspective, the full range of possible scores was observed at one or both assessments (data not shown). Score distributions were approximately symmetrical, with the exception of the body image scale at baseline which exhibited a positive skew (i.e. more patients scoring toward maximum levels of body image).

With respect to symptomatology, the full range of scores was also observed for the scales assessing micturition problems, male sexual problems and weight loss, at one or two assessment points. The weight loss scale was negatively skewed at both assessments (i.e. more patients scoring toward minimum levels of weight loss). The score distributions of the remaining symptom scales (i.e. gastrointestinal tract symptoms, chemotherapy side-effects, defaecation problems, stoma-related problems and female sexual problems) were generally restricted to the lower and upper middle part of the response scale at both assessments (i.e. higher and maximum scores were not observed).

Validity

Known-groups comparison. Time 1. As expected, patients with metastatic disease reported a significantly poorer future perspective ($P < 0.01$) and higher levels of gastrointestinal tract symptoms ($P < 0.05$) than patients with local or locoregional disease (Table 4). Surprisingly, patients with metastatic disease also reported significantly higher levels of chemotherapy side-effects. This unexpected finding could

Table 3. Sample size (n), means, standard deviations (S.D.) and internal consistency reliability (Cronbach's alpha) at baseline (T1) and time 2 (T2); sample size (n) and test–retest reliability (ICC) at times 3 and 4, of the QLQ-CR38

Scale/item	*n*	Mean	S.D.	Alpha	*n*	ICC (T3–T4)
Function						
BI T1	116	83.62	20.49	0.78	24	0.84
BI T2	108	81.93	20.96	0.79		
SX T1	97	23.26	24.95	0.86	18	0.92
SX T2	92	21.38	24.12	0.87		
SE T1	46	55.80	27.27	–	8	0.82
SE T2	41	50.41	29.93	–		
FU T1	117	55.84	29.31	–	23	0.53
FU T2	108	63.89	30.27	–		
Symptoms						
MI T1	117	26.12	20.75	0.63	23	0.79
MI T2	106	33.66	22.78	0.72		
GI T1	117	21.77	16.96	0.61	24	0.78
GI T2	108	27.04	16.90	0.49		
CT T1	117	11.21	15.12	0.47	24	0.90
CT T2	108	18.20	19.07	0.63		
DF T1	75	19.11	17.43	0.74	13	0.79
DF T2	72	27.01	19.09	0.79		
STO T1	43	22.68	17.66	0.76	8	0.83
STO T2	38	20.85	18.11	0.82		
MSX T1	71	41.79	40.35	0.83	9	0.79
MSX T2	63	42.94	39.66	0.91		
FSX T1	12	18.33	19.95	0.38	4	0.89
FSX T2	12	16.67	25.82	0.88		
WL T1	117	16.24	29.24	–	24	0.55
WL T2	107	10.59	19.75	–		

BI, body image; SX, sexual function; SE, sexual enjoyment; FU, future perspective; MI, micturition problems; GI, gastrointestinal tract symptoms; CT, chemotherapy side-effects; DF, defaecation problems; STO, stoma-related problems; MSX, male sexual problems; FSX, female sexual problems; WL, weight loss; ICC, intraclass correlation coefficient.

be attributed to high score levels of the item assessing dry mouth, a symptom that can also be caused by an absent or malfunctioning colon. The effect sizes of the significant group differences ranged from moderate to large (0.40 to 0.94).

Differences in pretreatment questionnaire scores as a function of performance status were in the expected direction. Patients with a poorer performance status (Karnofsky performance status of 70 or below) reported significantly higher levels of gastrointestinal tract symptoms ($P < 0.05$) and more weight loss ($P < 0.01$) than those with a better performance status (Table 4). Unexpectedly, patients with a poorer performance status reported higher levels of chemotherapy side-effects which, again, could be attributed to the item on dry mouth. The magnitude of the effect sizes was moderate to large (range 0.45–1.05).

Time 2. When on-treatment performance status was used as a grouping variable, statistically significant differences were in the expected direction for sexual functioning ($P < 0.01$), gastrointestinal tract symptoms ($P < 0.05$), chemotherapy side-effects ($P < 0.01$), defaecation problems ($P < 0.05$) and weight loss ($P < 0.01$) (Table 4). Effect sizes ranged from 0.44 to 0.82.

The presence of a stoma was also used as a grouping variable. According to expectation, patients with a stoma reported poorer body image ($P < 0.01$) than patients whose sphincters were intact and male patients with a stoma reported more sexual problems than those with preserved sphincters ($P < 0.05$). Additionally, patients with a stoma reported more micturition problems ($P < 0.05$) than patients with intact sphincters. This latter result may be attributed to either specific neurogenic bladder problems related to abdominoperineal resection and/or to the high prevalence of radiotherapy among stoma patients. The effect sizes were of a moderate magnitude (range 0.47–0.61).

Responsiveness. Using first and second assessment performance status (Karnofsky performance status) ratings as indicators of change in health status, the total patient sample was divided into those patients whose performance status had improved (17%), had remained unchanged (46%), or had deteriorated (37%). Statistically significant between-group differences over time (in ANOVA terms, group×time interactions) were observed for future perspective, weight loss ($P < 0.01$), micturition symptoms and chemotherapy side-effects ($P < 0.05$). In these cases, the change in mean scores was in the expected direction (data not presented in tabular form).

Repeated-measures ANOVA was used to test for between-group (radiotherapy versus chemotherapy) differences over time (first and second assessment) in scores on the QLQ-CR38 (Table 5). We expected patients who had received radiotherapy to report higher levels of micturition problems, gastrointestinal tract symptoms and defaecation problems at the second assessment than at baseline. Statistically significant between-group differences over time were observed in the expected direction for all three scales ($P = 0.0001$). Additionally, significant between-group differences over time were observed for body image ($P = 0.023$). Patients undergoing radiotherapy reported poorer and patients receiving chemotherapy reported better body image at the second assessment in comparison with baseline. Contrary to expectation, patients undergoing chemotherapy did not report significantly higher levels of chemotherapy-related side-effects over time.

Table 4. Summary of tests of known-groups comparisons (P values for between-group differences)

	Grouping variables*			
	Stage	KPS		Stoma
	Time 1	Time 1	Time 2	Time 2
Function				
BI	NS	NS	NS	$P<0.01$
SX	NS	NS	$P<0.01$	NS
FU	$P<0.01$	NS	NS	NS
SE	NS	NS	NS	NS
Symptoms				
MI	NS	NS	NS	$P<0.05$
CT	$P<0.05$	$P<0.01$	$P<0.01$	NS
GI	$P<0.05$	$P<0.05$	$P<0.05$	NS
MSX	NS	NS	NS	$P<0.05$
FSX	NS	NS	NS	NS
DF	NS	NS	$P<0.05$	
STO	NS	NS	NS	
WL	NS	$P<0.01$	$P<0.01$	NS

*Stage of disease: local/regional versus distant; Karnofsky performance status (KPS): 40–70 versus 80–100; Stoma: presence of a stoma versus intact sphincters. BI, body image; SX, sexual function; FU, future perspective; SE, sexual enjoyment; MI, micturition problems; CT, chemotherapy side-effects; GI, gastrointestinal tract symptoms; MSX, male sexual problems; FSX, female sexual problems; DF, defaecation problems; STO, stoma-related problems; WL, weight loss; NS, not significant.

Table 5. Responsiveness analysis comparing the change in scale scores of patient groups treated with radiotherapy or chemotherapy between baseline and the second assessment

			Mean (S.D.)			
Scale	Treatment group	*n*	Time 1	Time 2	*F*	*P* value
BI	RT	65	86.15 (19.25)	81.54 (21.36)	5.33	0.023
	CT	41	78.86 (22.33)	82.66 (20.79)		
MI	RT	64	28.30 (21.77)	42.53 (21.83)	17.53	0.0001
	CT	39	24.50 (19.94)	19.09 (15.91)		
CT	RT	66	9.26 (14.53)	16.16 (19.89)	–	NS
	CT	3	15.79 (16.26)	21.35 (17.40)	–	
GI	RT	67	21.39 (16.06)	31.64 (16.51)	21.14	0.0001
	CT	41	24.39 (17.55)	19.51 (14.86)		
DF	RT	33	20.92 (16.75)	39.39 (17.95)	28.64	0.0001
	CT	28	17.35 (15.78)	14.29 (11.22)		

BI, body image; MI, micturition problems; CT, chemotherapy side-effects; GI, gastrointestinal tract symptoms; DF, defaecation problems; RT, radiotherapy; CT, chemotherapy; S.D., standard deviation; NS, not significant. *F*, time×group interaction effect.

DISCUSSION

In this study we constructed a colorectal cancer-specific questionnaire module to be used in conjunction with the EORTC QLQ-C30. Module construction took place following the Study Group's guidelines. These included the identification of relevant QL issues, operationalisation of issues into questionnaire items and pretesting the provisional module. The resulting module consists of 38 items covering symptoms and side-effects related to different treatment modalities, stoma-related problems, body image, sexuality and future perspective. Because this module was devised to supplement the core instrument, a number of items are relevant for, but not entirely specific to colorectal cancer patients. These items pertain to chemotherapy side-effects, body image, sexuality, future perspective and weight loss, areas that are not or not sufficiently covered by the core instrument. Conversely, a number of items of the core instrument may be highly relevant for assessing the QL of colorectal cancer patients participating in clinical trials (e.g. items assessing constipation, diarrhoea, chemotherapy or radiotherapy side-effects). We therefore cannot recommend the QLQ-CR38 as a free-standing instrument for assessing the QL of colorectal cancer patients, but rather as a supplement to the core questionnaire.

To date, the cancer-specific QL questionnaires also used in clinical trials among cancer patients include the Rotterdam Symptom Checklist (RSCL) [28], the Cancer Rehabilitation Evaluation System Short Form (CARES-SF) [16], and the Functional Living Index-Cancer (FLIC) [29]. These questionnaires were designed to be applicable to a broad spectrum of cancer patients. The only other QL questionnaire that has been devised specifically for colorectal cancer patients is the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) [30]. This instrument was not in the public domain at the time we devised the QLQ-CR38. The FACT-C, like the QLQ-CR38, was designed for use among colorectal cancer patients with a range of disease stages and undergoing different treatments. In comparison with the QLQ-CR38, the FACT-C is shorter (36 versus 68 items, including the core instrument) and covers fewer symptoms and treatment-related side-effects.

In larger-scale testing among Dutch patients with colorectal cancer, the QLQ-CR38 was found to be feasible and was generally well accepted by the patients. On average, it required 10 min to complete both the QLQ-C30 and the QLQ-CR38 and, in most cases, it could be filled out by patients themselves, with little or no assistance. However, a number of male (3%) and female (12%) patients found the questions relating to sexuality too intrusive and, as a consequence, chose not to complete them. However, none of the patients refused to fill in the entire questionnaire. Similar results were found among patients with breast cancer who completed a breast cancer module—the EORTC QLQ-BR23—which also includes items assessing sexuality [8]. These findings confirm the general notion that missing data can be expected when sensitive areas are addressed in a questionnaire. However, at the same time, the results indicate that inquiring about sexuality need not be avoided as long as patients are left free to leave such questions unanswered.

With respect to the module's psychometric properties, the multitrait scaling analysis confirmed the hypothesised scale structure of the functioning scales, but did not strongly support the structure of the symptom scales. The item–rest correlations and the number of scaling successes of the symptom scales were, in general, relatively small. The chemotherapy side-effects scale merits special attention as this scale indicated consistent problems regarding both item internal consistency and item discriminant validity. Despite the inconsistent support for the scale structure of the QLQ-CR38 based on the multitrait scaling technique, seven of the nine multi-item scales met the minimal standards set for internal consistency reliability at one or both assessments. The two scales that did not meet these minimal requirements were those assessing symptoms of the gastrointestinal tract and chemotherapy side-effects. The test–retest reliability of all scales and one single item was found to be excellent ($ICC \geq 0.78$). However, the stability of two single items—future perspective and weight loss—was found to be lower (0.53 and 0.55, respectively). Since the mode of administration did not have a significant impact on the level of reported functioning or problems, with the exception of sexual functioning and male sexual problems, we are confident that the data on stability hold for the three administration procedures.

The validity of the QLQ-CR38 was indicated by its ability to discriminate between subgroups of patients known to differ in clinical status. According to expectation, selective scales distinguished clearly between patients differing in disease stage, initial and on-treatment performance status and the presence of a stoma. Additionally, selective scales were able to reflect changes in patients' Karnofsky performance status scores over time and were found to be responsive to treatment-induced change. The only unexpected finding that has not been explained was that the chemotherapy side-effects scale was not responsive to chemotherapy-induced change. This result may, in part, be attributed to the fact that the majority of patients received 5-FU and leucovorin, cytostatic drugs known for their limited toxicity. Despite its lack of responsiveness to chemotherapy-induced change, the chemotherapy side-effects scale was able to discriminate between patient subgroups and to reflect changes in patients' Karnofsky performance status scores over time.

It is not surprising that the symptom scales (particularly those assessing chemotherapy side-effects and gastrointestinal tract symptoms) are the weakest scales from a classic psychometric perspective. From a clinical point of view, the symptoms and side-effects involved are not necessarily expected to occur together. For example, the side-effects of chemotherapy—dry mouth, thin or lifeless hair and different taste—may not necessarily concur, and dry mouth may also result from intestinal malfunctioning. Additionally, the symptoms included in the gastrointestinal tract symptoms scale (see Appendix) may be induced by disease progression, surgical procedures and/or radiotherapy. However, even without evidence of a robust scale structure and high internal consistency reliability for scales assessing symptoms or side-effects, one can still combine those items that make sense on clinical grounds [31, 32]. The chemotherapy side-effects scale might also be disputed from a content validity perspective as it only contains three side-effects. Since these side-effects are very common, they can be considered as 'core' chemotherapy side-effects. They may need to be supplemented by additional, trial-specific chemotherapy side-effects to detect completely patients' experienced toxicity.

The relatively large number of scales and/or individual items generated by the core instrument and the colorectal cancer module combined has the advantage of increasing the level of informational detail, but unfortunately, also the number of outcome parameters. In hypothesis testing, this may result in multiple-testing problems where differences in mean scores between trial arms may achieve statistical levels of significance on the basis of chance alone. To limit the number of such outcome parameters, the EORTC Study Group on Quality of Life is currently developing higher order summary scores for the core instrument. Until such summary scores are available, we recommend that investigators identify, a priori, a limited set of the questionnaire subscales and single items considered to be of primary interest. The statistical analysis should then be focused primarily on this subset, whilst the remaining scales and single items can be analysed in a more exploratory manner.

Taken together, these results lend support to the clinical validity of the QLQ-CR38. Since the QLQ-CR38 provides detailed information on relevant aspects of colorectal cancer patients not covered by the QLQ-C30, it adds substantial information. The combination of the QLQ-C30 and the QLQ-CR38 will thus enhance the ability to detect clinically meaningful differences between treatment arms and clinically important changes in QL in colorectal cancer patients over time. However, an important limitation of the current investigation merits consideration. The EORTC Study Group on Quality of Life aims to develop modules for evaluating the QL of cancer patients participating in international clinical trials. The involvement of a range of languages in the construction and validation process simultaneously is, therefore, strongly recommended [5]. However, both the construction and validation of the colorectal cancer module have taken place in The Netherlands. Since this was one of the first module development projects taken on by the Study Group, each step was discussed extensively in the plenary sessions of the Study Group. The involvement of members representing a broad range of cultures and languages helped to ensure that the issues included were relevant across cultures and that the wording of the questionnaire items remained relatively simple and straightforward. Since many modules are currently being developed within the Study Group, extensive discussions in plenary sessions are time-prohibitive. The current, updated guidelines therefore require that module development takes

place in at least two languages and cultures simultaneously [6]. Whilst the module appears to perform well among Dutch colorectal cancer patients, its suitability needs to be tested in a range of other European languages. To date, the module is available in Danish, Dutch, English, German, French, Italian and Spanish. Finnish, Norwegian and Swedish translations are underway. These translations followed forward–backward translation procedures and pilot-testing among patients belonging to the target population, as detailed in the translation guidelines of the EORTC [33]. The QLQ-CR38 is currently being used in selective phase III clinical trials in different European countries. The data generated by these studies will provide the empirical basis for evaluating the module's performance when used in international cancer clinical trials.

1. Parkin DM, Muir CS, Whelen SL, *et al.* (eds). *Cancer Incidence in Five Continents*, Vol. VI. Lyon, IARC Scientific Publications, No. 120, 1992.
2. Sprangers MAG, Velde te A, Aaronson NK, Taal BG. Quality of life following surgery for colorectal cancer: a literature review. *Psycho-Oncol* 1993, **2**, 247–259.
3. Sprangers MAG. Quality of life assessment in colorectal cancer patients: evaluation of cancer therapies. *Semin Oncol* (in press).
4. Aaronson NK, Ahmedzai S, Bergman B, *et al.* for the European Organization for Research and Treatment of Cancer Study Group on Quality of Life. The EORTC QLQ-C30: a quality of life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993, **85**, 365–376.
5. Sprangers MAG, Cull A, Bjordal K, Groenvold M, Aaronson K. for the EORTC Study Group on Quality of Life. 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.
6. Sprangers MAG, Cull A, Groenvold M, Bjordal K, Blazeby J, Aaronson K. for the EORTC Quality of Life Study Group. The European Organization for Research and Treatment of Cancer approach to developing questionnaire modules: an update and overview. *Qual Life Res* 1998, **7**, 291–300.
7. Osoba D, Aaronson NK, Muller M, *et al.* The development and psychometric validation of a brain cancer quality of life questionnaire for use in combination with general cancer-specific questionnaires. *Qual Life Res* 1996, **5**, 139–150.
8. Sprangers MAG, Groenvold M, Arraras J, *et al.* on behalf of the EORTC Study Group on Quality of Life. The EORTC breast-cancer specific quality-of-life questionnaire (QLQ-BR23): first results from a three-country field study. *J Clin Oncol* 1996, **14**, 2756–2768.
9. Bjordal K, Ahlner-Elmqvist M, Tollesson E, *et al.* Development of a European Organisation 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.
10. Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. for the EORTC Study Group on Quality of Life. 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.
11. Blazeby JM, Alderson D, Winstone K, *et al.* on behalf of the EORTC Quality of Life Study Group. 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.
12. Howat SJ, Cull A, Greimel E, *et al. The Construction of an Ovarian Cancer Module.* Edinburgh, Internal report of the EORTC Study Group on Quality of Life, 1997.
13. Sprangers MAG, Cull A. *Guidelines for Module Development*. Amsterdam, Internal report of the EORTC Study Group on Quality of Life, 1992; revised version 1993.
14. Hopwood P. *Body Image*. Internal report of the EORTC Study Group on Quality of Life, Manchesster, U.K. 1992.
15. Sprangers MAG, Velde te A, Aaronson NK, *et al. The Construction of a Colorectal Cancer Module*. Amsterdam, Internal report of the EORTC Study Group on Quality of Life, 1992.
16. Schag CAC, Ganz PA, Heinrich RL. Cancer Rehabilitation Evaluation System-Short Form (CARES-SF): a cancer specific rehabilitation and quality of life instrument. *Cancer* 1991, **68**, 1406–1413.
17. Ware JE, Sherbourne CD. The MOS 36-item health survey (SF-36). Conceptual framework and item selection. *Med Care* 1992, **30**, 473–483.
18. Velde te A, Sprangers MAG, Aaronson NK. Feasibility, psychometric performance, and stability across modes of administration of the CARES-SF. *Ann Oncol* 1996, **7**, 381–390.
19. Aaronson NK, Muller M, Cohen PDA, *et al.* Translation, validation and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* (in press).
20. Schag CC, Heinrich RL, Ganz PA. Karnofsky Performance Status revisited: reliability, validity, and guidelines. *J Clin Oncol* 1984, **2**, 187–193.
21. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In MacLeod CM, ed. *Evaluation of Chemotherapeutic Agents*. New York, Columbia University Press, 1949, 191.
22. Hays RD, Hayashi T, Carson S, *et al. User's Guide for the Multitrait Analysis Program (MAP)*. Santa Monica, Rand Corporation, 1988.
23. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951, **16**, 297.
24. Nunnally JC, Bernstein IH. *Psychometric Theory*. New York, McGraw-Hill, 1994.
25. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966, **10**, 3.
26. Kerlinger FN. *Foundations of Behavioral Research*. New York, Holt, Rinehart & Winston, 1973.
27. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, New Jersey, Lawrence Erlbaum Associates, 1988.
28. 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.
29. Schipper H, Clinch J, McMurray A, Levitt M. Measuring the quality of life of cancer patients: the Functional Living-Index-Cancer: development and validation. *J Clin Oncol* 1984, **2**, 472–483.
30. Ward WL, Hahn EA, Mo F, *et al.* Reliability and validity of the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) quality of life instrument. *Qual Life Res* (in press).
31. Fayers PM, Hand DJ. Factor analysis, causal indicators and quality of life. *Qual Life Res* 1997, **6**, 139–150.
32. Fayers PM, Hand DJ, Bjordal K, Groenvold M. Causal indicators in quality of life research. *Qual Life Res* 1997, **6**, 393–406.
33. Cull A, Sprangers M, Aaronson NK. *EORTC Quality of Life Study Group Translations Procedure*. Edinburgh, Internal report of the EORTC Study Group on Quality of Life, 1994.

Acknowledgements—The authors thank Marc Bekkers, Henk Boot, Aafke Buitelaar, Ann Cull, Symone Detmar, Marianne van Eekeren, Hanneke de Haes, Penny Hopwood, Simon Horenblas, Thomas Küchler, Joos Lebesque, Anne Stiggelbout, Babs Taal, Jos Vroege, and Frans Zoetmulder for their contribution to the development of the colorectal cancer module; the members of the EORTC Study Group on Quality of Life for their useful comments on earlier versions of the colorectal cancer module; and Aafke Buitelaar, Liesbeth Abbink and Irma Kruyver for their contribution to the data collection. This study was supported by grants from the Dutch Cancer Society, Amsterdam, The Netherlands (NKI 91-30 and NKI 90-A). This research was conducted while Dr Sprangers was on the staff of the Netherlands Cancer Institute.

APPENDIX
CONTENT OF THE EORTC QLQ-CR38 AND ITS HYPOTHESISED SCALE STRUCTURE

Content area of items	Number of items
Function	
Body image (BI)	3
Feeling less attractive	
Feeling less feminine/masculine	
Dissatisfied with body	
Sexual functioning (SX)	2
Interest in sex	
Sexual activity	
Sexual enjoyment (SE)	1
Future perspective (FU)	1
Symptoms	
Micturition problems (MI)	3
Frequency of urination/day	
Frequency of urination/night	
Pain while urinating	
Symptoms in the area of the gastrointestinal tract (GI)	5
Bloated feeling in stomach	
Abdominal pain	
Pain in buttocks	
Bothered by gas (flatulence)	
Belching	
Chemotherapy side-effects (CT)	3
Dry mouth	
Thin or lifeless hair	
Different taste	
Problems with defaecation (only for patients with intact sphincters) (DF)	7
Frequency of bowel movements/day	
Frequency of bowel movements/night	
Urge without producing stools	
Unintentional release of stools	
Blood with stools	
Difficulty in moving bowels	
Painful bowel movements	
Stoma-related problems (only for patients with a stoma) (STO)	7
Afraid about stoma noise	
Afraid about smell of stools	
Worry about possible leakage	
Caring for stoma	
Irritated skin	
Embarrassment	
Feeling less complete	
Male sexual problems (only for men) (MSX)	2
Problems with erection	
Problems with ejaculation	
Female sexual problems (only for women who have been sexually active) (FSX)	2
Dry vagina	
Pain during intercourse	
Weight loss (WL)	1

One item not included in this overview is a screening question asking whether patients have a stoma. This item precedes the sections on defaecation and stoma-related problems.