

Keywords: cholangiocarcinoma; QLQ-BIL21; quality of life; gallbladder cancer; hepatobiliary cancer; PROM; EORTC

Validation of the EORTC QLQ-BIL21 questionnaire for measuring quality of life in patients with cholangiocarcinoma and cancer of the gallbladder

S D Kaupp-Roberts^{1,2}, G Yadegarfar³, E Friend¹, C M O'Donnell¹, J W Valle⁴, C Byrne⁵, I Bahar⁶, M Finch-Jones⁷, R Gillmore⁸, C D Johnson⁹, S P Pereira¹⁰, J K Wiggers¹¹, M Pinto¹², B Al-Sarireh¹³ and J K Ramage^{*1,14} on behalf of the EORTC Quality of Life Group

¹Department of Gastroenterology and Hepatology, Hampshire Hospitals NHS Foundation Trust, Aldermaston Road, Basingstoke RG24 9NA, UK; ²Faculty of Humanities and Social Sciences, University of Winchester, Sparkford Road, Winchester SO22 4NR, UK; ³School of Public Health, Isfahan University of Medical Sciences, Isfahan, Iran; ⁴Department of Medical Oncology, The Christie NHS Foundation Trust, Wilmslow Road, Withington, Manchester M20 4BX, UK; ⁵Hepatobiliary Cancer Services, Aintree University Hospital, Longmoor Lane, Liverpool, Merseyside L9 7AL, UK; ⁶Cachar Cancer Hospital and Research Centre, Meherpur, Silchar, Assam, India; ⁷Department of Surgery, University Hospitals Bristol NHS Foundation Trust, Upper Maudlin Street, Bristol BS2 8HW, UK; ⁸Department of Medical Oncology, Royal Free Hospital, Pond Street, London NW3 2QG, UK; ⁹Department of Surgery, University Hospital Southampton NHS Foundation Trust, Tremona Road, Southampton SO16 6YD, UK; ¹⁰UCL Institute for Liver and Digestive Health, Royal Free Hospital Campus, London NW3 2QG, UK; ¹¹Academic Medical Center, Meibergdreef 9, Amsterdam 1105 AZ, Netherlands; ¹²National Cancer Institute and G. Pascale Foundation of Naples, Naples, Italy; ¹³Morrison Hospital, ABM University Health Board, Heol Maes Eglwys, Morriston, Swansea SA6 6NL, UK and ¹⁴Institute of Liver Studies, King's College Hospital, Denmark Hill, London SE5 9RS, UK

Background: There is no specific quality of life (QoL) measurement tool to quantify QoL in patients with biliary tract cancer. Quality of life measurement is an increasingly crucial trial end point and is now being incorporated into clinical practice.

Methods: This International Multicentre Phase IV Validation Study assessed the QLQ-BIL21 module in 172 patients with cholangiocarcinoma and 91 patients with cancer of the gallbladder. Patients completed the questionnaire at baseline pretherapy and subsequently at 2 months. Following this, the psychometric properties of reliability, validity, scale structure and responsiveness to change were analysed.

Results: Analysis of the QLQ-BIL21 scales showed appropriate reliability with Cronbach's α -coefficients >0.70 for all scales overall. Intraclass correlations exceeded 0.80 for all scales. Convergent validity >0.40 was demonstrated for all items within scales, and discriminant validity was confirmed with values <0.70 for all scales compared with each other. Scale scores changed in accordance with Karnofsky performance status and in response to clinical change.

Conclusions: The QLQ-BIL21 is a valid tool for the assessment of QoL in patients with cholangiocarcinoma and cancer of the gallbladder.

*Correspondence: Professor JK Ramage; E-mail: john.ramage@hhft.nhs.uk

Received 17 April 2016; revised 10 July 2016; accepted 15 August 2016; published online 27 September 2016

© 2016 Cancer Research UK. All rights reserved 0007–0920/16

Malignant diseases of the biliary tree, encompassing cholangiocarcinoma and cancer of the gallbladder, are relatively rare, with evidence that incidence is rising (McGlynn *et al*, 2006). Incidence in the United States has been reported as between 1 and 2.5 cases per 100 000 per year (Kaushik, 2001; Khan *et al*, 2005), although rates of up to 96 per 100 000 have been reported in some high-risk populations (Shaib and El-Serag, 2004).

To date, cholangiocarcinoma and cancer of the gallbladder are associated with poor outcomes and impaired quality of life (QoL) (Heffernan *et al*, 2002). Patients presenting with advanced disease may have a mean survival of ≤ 12 months (Anderson *et al*, 2004). Extrahepatic disease has a marginally better prognosis than intrahepatic disease (Ahmed *et al*, 2008).

Presenting symptoms in cholangiocarcinoma are dependent on the location of the malignancy in the biliary tree (Anderson *et al*, 2004) and are directly related to obstruction of the biliary tree (DeOliveira *et al*, 2007). Intrahepatic disease may manifest as a liver mass or cholangitis with possible intrahepatic abscess formation as well as right upper quadrant pain related to capsular distension, whereas extrahepatic neoplasms lead to painless obstructive jaundice and cholangitis. Gallbladder cancer may be found incidentally at cholecystectomy; however, it may present with pain due to mass effects and may also trigger obstructive jaundice. In addition to local effects, systemic symptoms such as weight loss, fevers and anorexia are commonly seen.

Treatments consist of surgical resection (DeOliveira *et al*, 2007), biliary stenting or external drainage (Patel and Singh, 2007), chemotherapy (Hong *et al*, 2007) and radiotherapy (Ben-David *et al*, 2006). Significant survival benefits have not been reliably demonstrated (Malhi and Gores, 2006; Hong *et al*, 2007) and treatment-related QoL changes are poorly studied.

Quality of life and its measurements are increasingly regarded as crucial end points in clinical trials, as well as being gradually introduced into day-to-day clinical practice. There is paucity of data regarding QoL in cholangiocarcinoma and cancer of the gallbladder, and to date there is no disease-specific questionnaire available. The need for a disease-specific QoL questionnaire, and its subsequent development from phase I to III was first reported in 2011 (Friend *et al*, 2011). The European Organization for Research and Treatment of Cancer (EORTC) QLQ-BIL21 (Supplementary material) is a new disease-specific module to be used with the existing generic EORTC QLQ-C30 tool (Aaronson *et al*, 1993).

This study describes an International Phase IV Psychometric Validation Study designed to assess the clinical and psychometric reliability, validity and responsiveness to change of the QLQ-BIL21 questionnaire in patients with cholangiocarcinoma and cancer of the gallbladder.

MATERIALS AND METHODS

The EORTC guidelines (Blazeby *et al*, 2002) for module development were followed in this international multicentre study. The initial phases involve a process of collating a list of issues from patients, health-care workers and previous publications, which are subsequently developed into a bank of questions from which the module is constructed. The module is then tested in a Phase IV Validation Study.

Questionnaires. This study required patients to complete two paper questionnaires at all assessment time points: the well-established EORTC generic cancer questionnaires QLQ-C30 and the QLQ-BIL21. The QLQ-BIL21 consists of 21 questions: 3 single-item assessments relating to treatment side effects, difficulties with drainage bags/tubes and concerns regarding weight loss, in addition to 18 items grouped into 5 proposed scales: eating symptoms (4 items), jaundice symptoms (3 items), tiredness

(3 items), pain symptoms (4 items) and anxiety symptoms (4 items). The response format was a four-point Likert scale.

Responses to the QLQ-C30 and QLQ-BIL21 questionnaires were transformed into a 0–100 scale using EORTC guidelines. Translations according to EORTC guidelines (Dewolf *et al*, 2009) were completed for six languages (German, Dutch, Italian, Spanish, Mandarin Chinese and Hindi).

Recruitment. Patients were recruited between September 2011 and June 2014, and were eligible for inclusion if they were aged 18 years or above, with a diagnosis of cholangiocarcinoma or gallbladder cancer confirmed by histology or MDT (multidisciplinary team) opinion, were able to give written informed consent, had no prior history of other significant malignant disease, were able to understand the language of the questionnaire and had an expected minimum survival of 3 months. Patients were excluded if there was concurrent malignant disease elsewhere (except basal cell carcinoma of the skin), if prior surgery with curative intent had taken place (with no evidence of recurrence), or if psychological, familial, sociological or geographical reasons would hamper compliance with the study protocol. Ethics committee approval and individual written patient consent was obtained. The study protocol was approved by the EORTC QoL group.

Study design. QoL changes may be different following major surgical interventions when compared with procedural therapies or compared with supportive care alone; thus, patients were assigned to one of three groups depending on their planned treatment: Intervention Group 1 (surgical treatment including prior stents or drains), Intervention Group 2 (chemotherapy/radiotherapy/photodynamic therapy/laser therapy including prior stents or drains) and Intervention Group 3 (supportive care only, but not excluding stents or drains). Prior active treatment was not a barrier to participation provided all residual effects of the intervention had resolved by the time the patient was enrolled.

Patients completed both questionnaires (QLQ-C30 with QLQ-BIL21) at baseline. This was within 1 month before commencing treatment for Intervention Groups 1 and 2; for Intervention Group 3, the baseline assessment was performed as soon as the decision to proceed with supportive care had been made, and all patients completed a second set 2 months post baseline within a 2-week window. The Karnofsky performance status (KPS) was recorded at both time points. All patients completed a short paper debriefing questionnaire following the baseline assessment, covering issues such as completion time or identifying questions that were confusing, upsetting or difficult to answer.

Test–retest reliability was assessed in 67 clinically stable patients from across all three intervention groups 2 weeks after the 2-month follow-up assessment. Patients receiving intravenous chemotherapy at the time were excluded from participating in the test–retest element.

Statistical analysis

Reliability. The reliability, or internal consistency, of the QLQ-C30 and QLQ-BIL21 questionnaires was assessed by Cronbach's α -coefficient (Cronbach and Warrington, 1951). An internal consistency estimate of > 0.70 was considered acceptable for group comparison (Nunnally, 1994). Test–retest reliability was measured by calculating intraclass correlations (ICCs) for the QLQ-C30 and QLQ-BIL21 questionnaires. An ICC of > 0.90 is considered desirable, but values as low as 0.70 are acceptable for clinical use.

Validity. Multitrait scaling was used to examine the hypothesised scale structure of the individual items of the QLQ-BIL21. To test item scale convergence validity, a correlation of ≥ 0.40 was used. Comparison of an item with its own scale as compared with other

scales was used to support item-discriminant validity (Hays *et al*, 1988). Items were expected to correlate significantly better (at least twice the standard error) with its own scale than with other scales.

Three distinct approaches were used to evaluate the validity of the QLQ-BIL21:

- (i) *Convergent validity of the QLQ-C30 and QLQ-BIL21 questionnaires*: This was examined using Pearson's product-moment correlation. It was expected that conceptually related scales will correlate substantially with each other (Pearson's $r > 0.40$), and conversely scales that are less related will exhibit lower correlations (Pearson's $r < 0.40$).
- (ii) *Known group comparison of the QLQ-C30 and QLQ-BIL21 questionnaires*: This was performed to explore the ability of the questionnaire to discriminate between subgroups of patients differing in their clinical status, such as site (gallbladder *vs* intrahepatic cholangiocarcinoma *vs* extrahepatic cholangiocarcinoma *vs* disseminated disease) or treatment (resection possible *vs* inoperable). Differences were assessed using the *T*-test and if any were close to being nonsignificant, and were checked using Wilcoxon's rank sum for nonparametric data.
- (iii) *Responsiveness to clinical change over time of the QLQ-C30 and QLQ-BIL21 questionnaires*: This was assessed using the two sets of questionnaires available for each patient over time, with the mean QoL scores over time for items and scales reflecting change in QoL, the development of disease recurrence or metastases and performance status. These changes were compared by repeated-measure analysis of variance (ANOVA).

Statistical Package for the Social Science (SPSS) software (IBM Corporation, Armonk, NY, USA) was used for data analysis and the linear transformation of responses to the QLQ-C30 and QLQ-BIL21 questionnaires to a 0–100 scale using EORTC guidelines and SPSS syntax. Stata 11 was used for all analyses and a *P*-value of 0.01 was deemed significant. The method outlined in the EORTC scoring manual was used to impute missing data points (Fayers) <http://groups.eortc.be/qol/manuals>.

Sample size. The sample size was calculated to be 231 patients (Walter *et al*, 1998) based on the anticipation of a 5% attrition rate from enrolment to completion, and taking into account that a sample of 220 patients is required to achieve 80% power to detect differences in the α -coefficient under a null hypothesis of 0.60 and the alternative hypothesis of 0.70 using a two-sided *F*-test with a significance level of 0.05. A minimum sample of 210 is needed to achieve 10 responses per item (Tabachnik, 1993).

RESULTS

Enrolment. A total of 172 patients with cholangiocarcinoma and 91 patients with cancer of the gallbladder were recruited making a total of 263 eligible patients. Eleven centres in the United Kingdom (143 patients recruited overall), 2 in Germany (13 patients recruited overall) and 1 each in the Netherlands (14 patients), Italy (12 patients), Chile (4 patients), India (42 patients) and China (7 patients) enrolled a total of 263 patients fulfilling the study's inclusion criteria. Intervention Group 1 consisted of 44 patients receiving surgical treatments (16.9%), and Intervention Group 2 consisted of 103 patients receiving medical treatments (39.6%). Patients in the Intervention Group 3 (113 patients or 43.4%) received supportive care (no interventions with the exception of drains or stents). Three of the 263 patients were not attributed to a treatment group. Overall, 98 patients or 37.1% received a stent and drainage was performed in 31 or 11.7% of patients

Questionnaire completion. In total, 478 questionnaires were available for analysis. Of the 263 patients enrolled, 75 were not able to complete the study at 2 months (48 died, 11 were lost to follow-up, 13 too unwell and 3 for other reasons). Notably, 62% ($n = 30$) of patients who died were enrolled in India, representing 70% of all Indian patients.

Debriefing results. Of the 263 patients in the study, 256 completed the debriefing questionnaire at the baseline assessment. Of these, 66% ($n = 168$) did so in the hospital outpatient clinic, 10.9% at home ($n = 28$), 7.8% ($n = 20$) as an in-patient and 15.2% ($n = 39$) elsewhere. Overall, 89.1% of patients completed both QLQ-C30 and QLQ-BIL21 questionnaires in < 20 min, with 38.3% patients ($n = 98$) completing in < 10 min, 34.4% in 11–15 min ($n = 88$), 16.4% in 16–20 min ($n = 42$), 9% in 21–30 min ($n = 23$) and 1.6% taking over 30 min ($n = 4$). Help completing the questionnaire from family members or health-care staff was necessary in 46.4% of patients. A total of 41 patients (16%) found one or more items confusing or difficult to answer, the most common of which were question 49 (found confusing by 10 patients) and question 50 (found confusing by 4 patients) in the QLQ-BIL21 questionnaire. Forty-five per cent of patients ($n = 115$) found at least one question not relevant. In the QLQ-BIL21 questionnaire, question 50 (pertaining to drainage tubes and bags) was considered irrelevant by 29 patients. Question 49 was deemed irrelevant by 16 patients, and in the QLQ-BIL21 questionnaire, questions 37 and 35 (both pertaining to jaundice) were deemed irrelevant by 12 and 10 patients, respectively.

QLQ-BIL21 scale structure: consistency and reliability. The internal consistency of the QLQ-C30 and QLQ-BIL21 questionnaires was calculated for each scale at baseline, at 2-month follow-up and as an overall global score (Table 1). The overall global score of Cronbach's α coefficients was calculated using all data including at baseline as well as at 2 months. Cronbach's α -coefficients for all scales in the QLQ-BIL21 questionnaire exceeded the threshold of 0.70 at baseline (ranges 0.71–0.87). Overall Cronbach's α -coefficients for the QLQ-BIL21 scales all met the 0.70 threshold (ranges 0.71–0.89). For the QLQ-C30 questionnaire, a Cronbach's α -coefficient of > 0.70 was met for all scales except Physical Function ($\alpha = 0.47$), Cognitive Function ($\alpha = 0.65$) and Nausea/Vomiting ($\alpha = 0.67$) at baseline.

Intraclass correlations (Table 2) were calculated using the 67 individuals in the study who completed the test–retest questionnaire 2 weeks following the second assessment. ICC values for the QLQ-C30 scales varied from 0.52 to 0.92. All scales in the QLQ-BIL21 scale scored ICC values > 0.8 , the only ones scoring ≤ 0.9 being the Eating (ICC = 0.87), Jaundice (ICC = 0.86), Treatment Side effects (ICC = 0.83) and Weight Loss (ICC = 0.81). No ICC value in the QLQ-BIL21 scale fell below 0.80, well above the acceptable value of > 0.70 (Fayers, 2000).

Construct validity. Construct validity was determined for all scales using convergent and discriminant validity. Convergent validity was calculated for each scale in both the QLQ-C30 and QLQ-BIL21 questionnaires. For both questionnaires, all items had Pearson's *r* correlations > 0.4 for their own scales (Table 3).

Table 1. Cronbach's α for BIL21 scales

Scale	Cronbach's α Baseline	Cronbach's α 2 months	Cronbach's α Overall
BIL21			
Eating	0.75	0.77	0.75
Jaundice	0.71	0.61	0.71
Tiredness	0.87	0.93	0.89
Pain	0.81	0.68	0.78
Anxiety	0.8	0.8	0.81

Table 2. Results for test-retest for BIL21 scales

Scale	n	R ²	ICC
ICC for QLQ-BIL21 scales			
Eating	67	0.9403	0.875
Jaundice	66	0.9373	0.8665
Tiredness	67	0.9598	0.9157
Pain	66	0.9724	0.9431
Anxiety	67	0.9802	0.9585
Treatment side effects	67	0.9172	0.8268
Drains	66	0.9598	0.9143
Weight loss	67	0.9092	0.8101

This confirms the clinical reasoning that the individual questions have been placed in the correct scales and are broadly measuring aspects of the same theme. Discriminant validity was calculated for all items in the QLQ-BIL21 questionnaire, none of which were >0.70. This indicates that none of the items correlate with scales outside the scale they have been placed into and indicated that items must therefore be in the correct scale.

Clinical validity. The ability of the QLQ-BIL21 questionnaire to correlate with clinical scores of function at discrete time points was assessed using KPS scores at baseline stratified into two groups (KPS <70 and >70). The difference in scores for each scale in these two groups subsequently underwent group comparison testing (n = 238–256; Table 4). This shows that the QLQ-BIL21 questionnaire is able to identify significant differences in mean scale scores between the two Karnofsky groups in all scales except Jaundice (P = 0.139) and Weight loss (P = 0.898). This may be explained by the fact that weight loss is a less prominent feature in biliary malignancies than in other GI cancers. The lack of correlation between jaundice and low KPS is of particular interest.

Responsiveness to clinical change over time. Clinical change over time was assessed by comparing mean scale scores at baseline and follow-up in 154–178 patients (this variation was due to some scales not being fully completed at the follow-up assessment) (Table 5). Changes in mean scores by scale over time are a measure of clinical responsiveness. These changes were only significant for eating, jaundice, tiredness, pain and treatment side effects. A change in mean scale score of >10% was only seen in the treatment side effects item (Δ12.54). Changes over the 2-month follow-up period for anxiety (P = 0.066), drains (P = 0.834) and weight loss (P = 0.519) were not significant.

Analysis of the changes to scales over time stratified by the intervention group (Table 5) showed more subtle changes. Eating symptoms changed minimally by a nonsignificant margin in all three intervention groups. Jaundice symptoms showed a modest improvement (change of –13.7) in the Intervention Group 1 (surgery), with a smaller change (–5.6) in the Intervention Group 2 (medical therapies including chemotherapy and radiotherapy). There were no significant changes to jaundice symptoms in the Intervention Group 3 (supportive care only but including biliary stents and drains). Tiredness symptoms worsened significantly in the Intervention Groups 1 (+12.0) and 3 (+8.8), with insignificant change in the Intervention Group 2. Pain and anxiety symptoms improved significantly (change of –11.5 and –7.9, respectively) in the Intervention Group 2, with no significant changes in either Intervention Group 1 or 3. Treatment side effects worsened markedly in the Intervention Group 2 (+19.7), with no significant changes in Intervention Groups 1 and 3. Symptoms regarding drains and weight loss did not change significantly across all three intervention groups between baseline and follow-up. The worsening of weight loss symptoms (+10.0) in the Intervention Group 3 was not statistically significant, but had a significant clinical change.

Table 3. Construct validity for QLQ-BIL21 questions

Correlation with Eating scale	
1	0.843
2	0.789
3	0.751
4	0.71
Correlation with Jaundice scale	
5	0.896
6	0.759
7	0.812
Correlation with Jaundice scale	
8	0.91
9	0.937
10	0.919
Correlation with Pain scale	
11	0.842
12	0.863
13	0.799
14	0.759
Correlation with Anxiety scale	
15	0.789
16	0.761
17	0.843
18	0.789
Correlation with single items	
19	1
20	1
21	1

Table 4. Known-group comparison: correlation with KPI

BIL21 scale	KPI <70 Mean score	KPI >70 Mean score	Difference Δ	Significance (P-values)
Eating	50.18	32.23	17.98	<0.01
Jaundice	30.19	22.75	7.44	0.139
Tiredness	74.32	52.69	21.62	<0.01
Pain	48.52	33.05	15.46	0.01
Anxiety	63.22	49.25	13.97	0.01
Treatment side effects	43.65	28.02	15.63	0.05
Drains	28.68	10.32	18.36	0.01
Weight loss	35.51	34.755	0.75	0.898

Abbreviation: KPI = Karnofsky performance index. Bold indicates non-significance.

Differences between cholangiocarcinoma and cancer of the gallbladder. As this module covers more than one group of diseases clinically, differences between these groups were calculated. It might be expected that jaundice may be more prominent in cholangiocarcinoma than gallbladder cancer, but there was no clear difference in this scale. Mean scores for the QLQ-BIL21 were calculated at baseline and at follow-up, and any group differences between cholangiocarcinoma and cancer of the gallbladder were investigated using repeated-measure ANOVA. Data for the baseline QLQ-BIL21 are displayed in Table 6. There was a significant difference between extra- and intrahepatic cholangiocarcinoma in the drain sites and weight loss questions only, as might be expected clinically since drains would be used more for intrahepatic than extrahepatic disease, and weight loss might be expected if there is a significant bulk of intrahepatic disease. None of the other scales were significantly different at baseline.

Table 5. Change over time for whole group and intervention subgroups 1–3

Scale/items	No.	Range of change	Mean change	s.d.	Median	Significance P-values
Overall						
Eating scale	177	– 66.67 to 83.33	3.71	24.7	0.0	0.047
Jaundice scale	177	– 100 to 77.78	– 7.65	29.0	0.0	0.001
Tiredness scale	178	– 66.67 to 100	6.05	31.1	0.0	0.01
Pain scale	175	– 83.33 to 33.33	– 6.57	21.2	0.0	<0.01
Anxiety scale	176	– 58.33 to 58.33	– 3.3	23.5	0.0	0.066
Treatment side effects	170	– 100 to 100	12.5	38.4	0.0	<0.01
Drains	154	– 100 to 100	– 0.43	25.6	0.0	0.834
Weight loss	175	– 100 to 100	1.71	35.1	0.0	0.519
Scale/items	No.	Mean change	s.d.	Median	Significance P-values	
Intervention Group 1						
Eating scale	37	4.28	23.4	0.0		0.281
Jaundice scale	38	– 13.74	30.2	0.0		0.008
Tiredness scale	37	12.01	30.1	11.1		0.02
Pain scale	38	– 1.97	23.0	0.0		0.601
Anxiety scale	37	– 4.28	25.0	– 8.3		0.304
Treatment side effects	37	4.50	37.8	0.0		0.473
Drains	36	5.56	28.2	0.0		0.245
Weight loss	36	0.93	0.84	0.0		0.845
Intervention Group 2						
Eating scale	86	5.23	25.5	8.3		0.06
Jaundice scale	87	– 5.62	24.0	0.0		0.031
Tiredness scale	87	1.79	31.1	0.0		0.593
Pain scale	85	– 11.57	21.1	– 8.3		<0.01
Anxiety scale	85	– 7.94	20.7	– 8.3		<0.01
Treatment side effects	81	19.75	37.2	33.3		<0.01
Drains	73	– 4.57	21.0	0.0		0.068
Weight loss	86	– 3.10	33.0	0.0		0.386
Intervention Group 3						
Eating scale	54	0.93	24.2	0.0		0.78
Jaundice scale	52	– 6.62	35.1	0.0		0.179
Tiredness scale	54	8.85	31.5	0.0		0.044
Pain scale	52	– 1.76	18.5	0.0		0.496
Anxiety scale	54	4.78	24.7	8.3		0.161
Treatment side effects	52	7.05	39.2	0.0		0.201
Drains	45	1.48	29.3	0.0		0.736
Weight loss	53	10.06	41.1	0.0		0.081

This shows reduction in jaundice in groups 1 and 2 (due to interventions). In group 2, a reduction in pain and anxiety scores was seen owing to intervention but with an increase in treatment side effects. Bold indicates non-significance.

DISCUSSION

The current study is the final stage of a process of development and validation of a specific set of questions appropriate for biliary tract cancers. The QLQ-BIL21 questionnaire was conceived as the first disease-specific QoL assessment tool for patients with cholangiocarcinoma and cancer of the gall bladder. Although the existing EORTC questionnaire modules QLQ-PAN26 (Pancreatic Cancer Module, presently awaiting phase IV validation) (Fitzsimmons *et al*, 1999), the QLQ-LMC21 (Colorectal Liver Metastasis module) (Blazeby *et al*, 2009) and the FACIT.org questionnaire FACT-Hep (for all hepatobiliary cancers) covered many similar domains, and indeed some of their issues contributed to the construction of the QLQ-BIL21, the differences in presentation and disease course are reflected in a significantly different scale structure.

This International Phase IV Validation Study of the QLQ-BIL21 questionnaire conducted in accordance with EORTC guidelines found the reliability and validity of the questionnaire to be robust. Cronbach's α -coefficients for all scales overall met the threshold of >0.70. The robustness of the testing parameters indicate that the QLQ-BIL21 questionnaire will prove clinically useful in trials as well as in clinical practice. The reliability analysis of the questionnaire did not fully meet the ideal ICC threshold of >0.90 for all scales (the exceptions being: eating, jaundice, treatment side effects and weight loss), but exceeded the minimum reported standard of 0.70. The QLQ-C30 questionnaire behaved

differently in this population, with two scales failing to score an ICC >0.70. These results are an indication that the QLQ-BIL21 questionnaire may be needed in addition to the QLQ-C30 questionnaire in this patient population.

Analysis of the debriefing questionnaires highlighted that a large number of patients found QLQ-BIL21 question 50 ('Have you had difficulties with drainage tubes/bags) irrelevant. This is not surprising as not all patients will encounter drains in the course of their treatment. To maintain clarity, it may therefore be desirable to enable patients to answer this question and also question 49 with 'not applicable' rather than a Likert scale in the final version of the questionnaire.

Patients enrolled in India differed statistically from those in other centres for some characteristics. The Indian cohort of 43 patients was significantly younger with a mean age of 52.24 years, a decade below the next youngest cohort in the Netherlands (mean age 62.6 years) and over two decades younger than the oldest cohort in the United Kingdom (Winchester mean age 75.5 years). Furthermore, 93% of Indian patients had primary gallbladder cancer, compared with an aggregated mean of 21.7% for all other centres. Overall, 43% of all gallbladder cancer patients enrolled in the study came from the Indian cohort. Mortality was also significantly higher in the Indian cohort, with a mortality of 70% ($n = 30$) compared with an overall mortality of 18%, whereas excluding the Indian patients from the analysis yields an overall mortality of just 8%. Despite this demographic background, there seems to be no significant differences in QLQ-BIL21 scores for any

Table 6. Mean baseline scores for individual sites of disease showing *n*, mean and *s.d.* for extrahepatic, intrahepatic and gallbladder cancer

Scale	No.	Mean score at baseline	s.d.
Eating scale			
Extrahepatic cholangiocarcinoma	78	32.58	27.2
Intrahepatic cholangiocarcinoma	73	32.42	25.6
Gallbladder cancer	87	39.94	28.2
Jaundice scale			
Extrahepatic cholangiocarcinoma	78	18.37	25.8
Intrahepatic cholangiocarcinoma	74	25.97	28.2
Gallbladder cancer	85	30.32	31.8
Tiredness scale			
Extrahepatic cholangiocarcinoma	78	58.11	31.3
Intrahepatic cholangiocarcinoma	74	55.55	31.4
Gallbladder cancer	83	55.28	31.6
Pain scale			
Extrahepatic cholangiocarcinoma	78	30.12	25.0
Intrahepatic cholangiocarcinoma	73	34.36	26.4
Gallbladder cancer	85	42.15	31.8
Anxiety scale			
Extrahepatic cholangiocarcinoma	78	52.67	26.6
Intrahepatic cholangiocarcinoma	71	48.98	26.4
Gallbladder cancer	84	52.9	33.3
Treatment side effects			
Extrahepatic cholangiocarcinoma	75	31.55	32.4
Intrahepatic cholangiocarcinoma	71	26.76	30.1
Gallbladder cancer	84	35.31	32.3
Drains			
Extrahepatic cholangiocarcinoma	73	5.93	17.0
Intrahepatic cholangiocarcinoma	70	13.33	26.0
Gallbladder cancer	82	21.13	33.3
Weight loss			
Extrahepatic cholangiocarcinoma	78	25.64	28.4
Intrahepatic cholangiocarcinoma	73	43.83	38.0
Gallbladder cancer	86	34.88	35.0

Bold shows the only significant differences (both $P=0.01$).

scale, with the exception of jaundice, which is broadly congruent with clinical expectation.

Known group comparison testing demonstrated the QLQ-BIL21 questionnaire to be adept at identifying group differences at time points. Of particular interest here was the low correlation between jaundice and a low KPS. This may be a reflection of the fact that even a very small tumour load may generate jaundice without a concurrent effect on systemic symptoms or possibly that jaundice is often treated at an early stage while the tumour progresses. Similar results were found in the development of the QLQ-PAN26, although the jaundice items were retained on expert opinion.

Overall, the measured response to clinical change was not as marked as anticipated, but was significantly different from baseline in most scales. This may be a reflection of the short follow-up period, which was initially selected to prevent patient drop out owing to clinical deterioration and because it represents the period of maximal clinical intervention. It would have been desirable to continue data collection beyond this; however, the poor prognosis of this rare cancer would have led to a highly significant attrition rate. Indeed, even at 2 months 50 out of the 263 patients recruited had died. Comparable studies include a 2010 study of 91 patients in Romania who were followed up for up to 4 years with regular QLQ-C30 assessments (Mihalache *et al*, 2010). At the first 6-month follow-up appointment, statistically significant increases in global scores were detected. Conversely, a 2015 Thai study of

99 patients found significant decreases in QoL using the FACT-Hep Questionnaire at 2 months (Woradet *et al*, 2015). It may therefore prove desirable to further investigate the QLQ-BIL21 questionnaire in a smaller cohort over a longer time period to assess response to change over time more thoroughly. When each group was analysed separately, the changes identified in each treatment group by the QLQ-BIL21 questionnaire are broadly congruent with clinical expectation, with less change in Group 3 compared with the other active treatment groups.

Overall, the strengths of this study are that it followed EORTC module development guidelines, which included many different languages and cultures, including patients with a wide variety of stages, therapies and presentations of this disease. A drawback of this study is that a large proportion of patients were recruited from the United Kingdom, which may have caused some biases in terms of language and culture, although it is debatable whether the results would have been materially different had a more even spread of international patients been recruited.

Changes to biliary tract cancer therapeutics are evolving. The first advanced biliary tract cancer randomised phase II trial (ABC-01) of gemcitabine vs cisplatin/gemcitabine was reported in 2009 (Valle *et al*, 2009), leading to the pivotal randomised phase III study (ABC-02) establishing cisplatin/gemcitabine chemotherapy as the international standard of care regimen for patients with these disease (Valle *et al*, 2010). Its sixth iteration (ABC-06), a phase III randomised controlled trial of oxaliplatin and 5-FU vs active symptom control in the second-line setting after failure of cisplatin/gemcitabine chemotherapy, is currently active with recruitment closing in summer 2016; this study is one of the first to incorporate the use of QLQ-BIL21. The changes in therapeutics occurring as a result of these trials may require either a future adaptation of the QLQ-BIL21 questionnaire to include side effects of new drugs or an additional module specifically for new therapies.

CONCLUSION

The QLQ-BIL21 questionnaire has been demonstrated to be a clinically sensitive, reliable and valid instrument for measuring QoL in patients with cholangiocarcinoma and cancer of the gallbladder by this International Multicentre Phase IV Validation Study. The authors therefore recommend that this tool be used in the clinical trial and clinical practice setting to provide an accurate quantification of QoL to guide therapy and future research.

ACKNOWLEDGEMENTS

We thank all centres, investigators and patients who took the time to participate in the study. Particular gratitude is extended to the following collaborators, without whom this study would not have been possible: Dr Serena Hilman, Weston Area Health Trust, UK; Dr Julia Quidde, University Hospital Hamburg-Eppendorf, Hamburg, Germany; Prof Jörg Schirra, Munich, Germany; Dr Maria Elisa Irrarázaval, Santiago, Chile; Dr Bao Junjun, First Affiliated Hospital of Anhui Medical University, China; Dr Heinz-Joseph Klumpen, Academic Medical Center, Amsterdam, The Netherlands. This project was financially supported by a grant from the EORTC Quality of Life Group. The project received service support from the National Institute of Health Research, United Kingdom.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, De Haes CJM, Kaasa S, Klee M, Osoba D, Razavi D, Rofe B, Schraub S, Sneeuw K, Sullivan M, Takeda F (1993) 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* **85**: 365–376.
- Ahmed K, Beg MS, Kromokji R, Safa MM (2008) Intrahepatic versus extrahepatic cholangiocarcinoma: is there a difference in survival? *ASCO Meet Abstr* **26**: 15564.
- Anderson CD, Pinson CW, Berlin J, Chari RS (2004) Diagnosis and treatment of cholangiocarcinoma. *Oncologist* **9**: 43–57.
- Ben-David MA, Griffith KA, Abu-Isa E, Lawrence TS, Knol J, Zalupski M, Ben-Josef E (2006) External-beam radiotherapy for localized extrahepatic cholangiocarcinoma. *Int J Radiat Oncol Biol Phys* **66**: 772–779.
- Blazeby J, Sprangers M, Cull A, Groenvold M, Bottomley A (2011) *EORTC Quality of Life Group Guidelines for Developing Questionnaire Modules (EORTC Quality of Life Group)* 4th edition, Brussels: EORTC Quality of Life study Group.
- Blazeby JM, Fayers P, Conroy T, Sezer O, Ramage J, Rees M (2009) Validation of the European Organization for Research and Treatment of Cancer QLQ-LMC21 questionnaire for assessment of patient-reported outcomes during treatment of colorectal liver metastases. *Br J Surg* **96**: 291–298.
- Cronbach LJ, Warrington WG (1951) Time-limit tests: estimating their reliability and degree of speeding. *Psychometrika* **16**: 167–188.
- DeOliveira ML, Cunningham SC, Cameron JL, Kamangar F, Winter JM, Lillemoe KD, Choti MA, Yeo CJ, Schulick RD (2007) Cholangiocarcinoma: thirty-one-year experience with 564 patients at a single institution. *Ann Surg* **245**: 755–762.
- DeWolf L, Koller M, Velikova G, Johnson C, Scott N, Bottomley A (2009) *EORTC Quality of Life Group Translation Procedure*, 3rd Edition. EORTC Quality of life group: Brussels, Belgium.
- Fayers PM (2000) *Quality of Life: Assessment, Analysis and Interpretation*. Wiley: New York, NY, USA.
- Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A (2001) *The EORTC QLQ-C30 Scoring Manual*, 3rd edn. EORTC Quality of life group: Brussels, Belgium.
- Fitzsimmons D, Johnson CD, George S, Payne S, Sandberg AA, Bassi C, Beger HG, Birk D, Büchler MW, Dervenis C, Fernandez Cruz L, Friess H, Grahm AL, Jeekel J, Laugier R, Meyer D, Singer MW, Tihanyi T (1999) Development of a disease specific quality of life (QoL) questionnaire module to supplement the EORTC core cancer QoL questionnaire, the QLQ-C30 in patients with pancreatic cancer. EORTC Study Group on Quality of Life. *Eur J Cancer* **35**: 939–941.
- Friend E, Yadegarfar G, Byrne C, Johnson CD, Sezer O, Pucciarelli S, Pereira SP, Chie W-C, Banfield A, Ramage JK (2011) Development of a questionnaire (EORTC module) to measure quality of life in patients with cholangiocarcinoma and gallbladder cancer, the EORTC QLQ-BIL21. *Br J Cancer* **104**: 587–592.
- Hays RD, Havashi T, Carson S, Ware JE (1988) *User's Guide for the Multitrait Analysis Program (MAP)*. The RAND Corporation: Santa Monica, CA, USA.
- Heffernan N, Cella D, Webster K, Odom L, Martone M, Passik S, Bookbinder M, Fong Y, Jarnagin W, Blumgart L (2002) Measuring health-related quality of life in patients with hepatobiliary cancers: the functional assessment of cancer therapy-hepatobiliary questionnaire. *J Clin Oncol* **20**: 2229–2239.
- Hong YS, Lee J, Lee SC, Hwang IG, Choi S-H, Heo J-S, Park JO, Park YS, Lim HY, Kang WK (2007) Phase II study of capecitabine and cisplatin in previously untreated advanced biliary tract cancer. *Cancer Chemother Pharmacol* **60**: 321–328.
- Kaushik SP (2001) Current perspectives in gallbladder carcinoma. *J Gastroenterol Hepatol* **16**: 848–854.
- Khan SA, Thomas HC, Davidson BR, Taylor-Robinson SD (2005) Cholangiocarcinoma. *Lancet (London, England)* **366**: 1303–1314.
- Malhi H, Gores GJ (2006) Cholangiocarcinoma: modern advances in understanding a deadly old disease. *J Hepatol* **45**: 856–867.
- McGlynn KA, Tarone RE, El-Serag HB (2006) A comparison of trends in the incidence of hepatocellular carcinoma and intrahepatic cholangiocarcinoma in the United States. *Cancer Epidemiol Biomarkers Prev* **15**: 1198–1203.
- Mihalache F, Tantau M, Diaconu B, Acalovschi M (2010) Survival and quality of life of cholangiocarcinoma patients: a prospective study over a 4 year period. *J Gastrointest Liver Dis* **19**: 285–290.
- Nunnally J, Bernstein IH (1994) *Psychometric Theory*, 3rd edn. McGrawHill: New York, NY, USA.
- Patel T, Singh P (2007) Cholangiocarcinoma: emerging approaches to a challenging cancer. *Curr Opin Gastroenterol* **23**: 317–323.
- Shaib Y, El-Serag HB (2004) The epidemiology of cholangiocarcinoma. *Semin Liver Dis* **24**: 115–125.
- Tabachnik BJ FL (1993) *Using Multivariate Statistics*. Harper and Row: London, UK.
- Valle J, Wasan H, Palmer DH, Cunningham D, Anthony A, Maraveyas A, Madhusudan S, Iveson T, Hughes S, Pereira SP, Roughton M, Bridgewater J, ABC-02 Trial Investigators (2010) Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. *N Engl J Med* **362**: 1273–1281.
- Valle JW, Wasan H, Johnson P, Jones E, Dixon L, Swindell R, Baka S, Maraveyas A, Corrie P, Falk S, Gollins S, Lofts F, Evans L, Meyer T, Anthony A, Iveson T, Highley M, Osborne R, Bridgewater J (2009) Gemcitabine alone or in combination with cisplatin in patients with advanced or metastatic cholangiocarcinomas or other biliary tract tumours: a multicentre randomised phase II study – The UK ABC-01 Study. *Br J Cancer* **101**: 621–627.
- Walter SD, Eliasziw M, Donner A (1998) Sample size and optimal designs for reliability studies. *Stat Med* **17**: 101–110.
- Woradet S, Promthet S, Songserm N, Parkin DM (2015) Factors affecting health-related quality of life in patients with cholangiocarcinoma in the northeastern region of Thailand *Cancer Nurs* **38**: E46–E51.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)