EUROPEAN UROLOGY 66 (2014) 1148–1156

available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Bladder Cancer

Validation and Reliability Testing of the EORTC QLQ-NMIBC24 Questionnaire Module to Assess Patient-reported Outcomes in Non–Muscle-invasive Bladder Cancer

Jane M. Blazeby[a,b,*], *Emma Hall*[c], *Neil K. Aaronson*[d], *Lisa Lloyd*[c], *Rachel Waters*[c], *John D. Kelly*[e], *Peter Fayers*[f,g]

[a] *Centre for Surgical Research, School of Social and Community Medicine, Bristol, UK;* [b] *Division of Surgery, Head and Neck, University Hospitals NHS Foundation Trust, Bristol, UK;* [c] *Institute of Cancer Research Clinical Trials and Statistics Unit, London, UK;* [d] *Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands;* [e] *Division of Surgery and Interventional Science, University College London, London, UK;* [f] *Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK;* [g] *Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway*

Article info

Article history:
Accepted February 13, 2014
Published online ahead of print on February 25, 2014

Keywords:
Bladder cancer
Quality of life
EORTC questionnaire
Validation

Abstract

Background: Well-developed and well-tested patient-reported outcome measures for non–muscle-invasive bladder cancer (NMIBC) are required.
Objective: To test and adapt the scale structure and explore the psychometric properties of the European Organisation for Research and Treatment of Cancer (EORTC) questionnaire for NMIBC.
Design, setting, and participants: A total of 433 patients in the Bladder COX-2 Inhibition Trial (BOXIT) completed the EORTC QLQ-C30 and NMIBC questionnaires. BOXIT is evaluating the addition of celecoxib to standard treatment in high- and intermediate-risk NMIBC.
Outcome measurements and statistical analysis: Multitrait scaling investigated and adapted the questionnaire scale structure and evaluated the reliability and validity of the revised scales, as well as responsiveness to change.
Results and limitations: A total of 410 patients (94.7%) (79.3% men, 74.6% high risk) returned baseline forms, and the questionnaire response rate was 88.2%. Multitrait scaling confirmed six scales and five single items. Scales and items demonstrated significant differences between patients with good and poor performance status scores ($p < 0.001$). Men reported better sexual function than women ($p < 0.001$). Scale and single-item module scores were not highly correlated with QLQ-C30 scores (evidence of discriminant validity), and the module was responsive to changes in health over time. International and test–retest data are required.
Conclusions: This study demonstrates the evidence-driven adapted scale structure and psychometric data of the EORTC QLQ-NMIBC24 module to use in clinical trials of patients with high- or intermediate-risk bladder cancer.

* Corresponding author. University Hospitals NHS Foundation Trust, Bristol Royal Infirmary, Upper Maudlin Street, Bristol, Avon BS2 8HW, UK.
E-mail address: j.m.blazeby@bris.ac.uk (J.M. Blazeby).

http://dx.doi.org/10.1016/j.eururo.2014.02.034
0302-2838

1. Introduction

The majority of patients with bladder cancer (BCa) present with non–muscle-invasive BCa (NMIBC) and are managed by endoscopic resection alone plus immediate postoperative intravesical chemotherapy [1]. Depending on risk stratification, intravesical immunotherapy with bacillus Calmette-Guérin (BCG) or chemotherapy using mitomycin C (MMC) may be considered. Evaluation of current treatments today typically includes assessment of patient-reported outcomes (PROs) in addition to clinical end points. PROs are defined as outcomes from the patients themselves that are not interpreted by an observer [2]. Measurement of PROs is most commonly undertaken with questionnaires, and the European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 and the Functional Assessment of Cancer Therapy measures are widely used. Both assess generic aspects of health and symptoms that commonly occur with cancer [3–6]. Measures may be supplemented by disease-specific modules to address concerns in specific cancer sites.

In the 1990s, the EORTC Quality of Life Group developed modules for BCa, the QLQ-BLS24 for superficial BCa (NMIBC) and the QLQ-BLM30 for muscle-invasive BCa [7]. Both modules have been used in clinical studies, but formal validation data are lacking. The aim of this study was to examine the scale structure, reliability, and clinical validity of the QLQ-BLS24 in patients with NMIBC.

2. Methods

Patients participating in the Bladder COX-2 Inhibition Trial (BOXIT; CR UK/07/004; ISRCTN: 84681538) were recruited. BOXIT is a randomised placebo-controlled trial evaluating the addition of celecoxib to standard treatment (transurethral resection of bladder tumour, single-dose MMC, and BCG induction and maintenance for disease at high risk for recurrence or multiple MMC instillations for disease at intermediate risk for recurrence [8]). Patients with primary or recurrent NMIBC at high or intermediate risk of recurrence according to the 2002 European Association of Urology guidelines were eligible and include Tis, T1, and Ta tumours other than those at low risk [9]. The interventions in BOXIT were administered according to the study protocol [8].

2.1. Questionnaires

Patients completed the QLQ-C30 questionnaire and the QLQ-BLS24 module before treatment in a clinic and at regular intervals thereafter. In the high-risk groups, questionnaires were completed at time point 0 and at 2, 3, 6, and 12 mo. In the intermediate-risk group, assessments were completed at time point 0 (before randomisation) and at 12 mo (Fig. 1). Missing data were imputed according to the EORTC guidelines, and questionnaires were considered as missing if >50% of the items were missing [4]. Using this approach, some missing items still could not be imputed, but the other data from these questionnaires were still used.

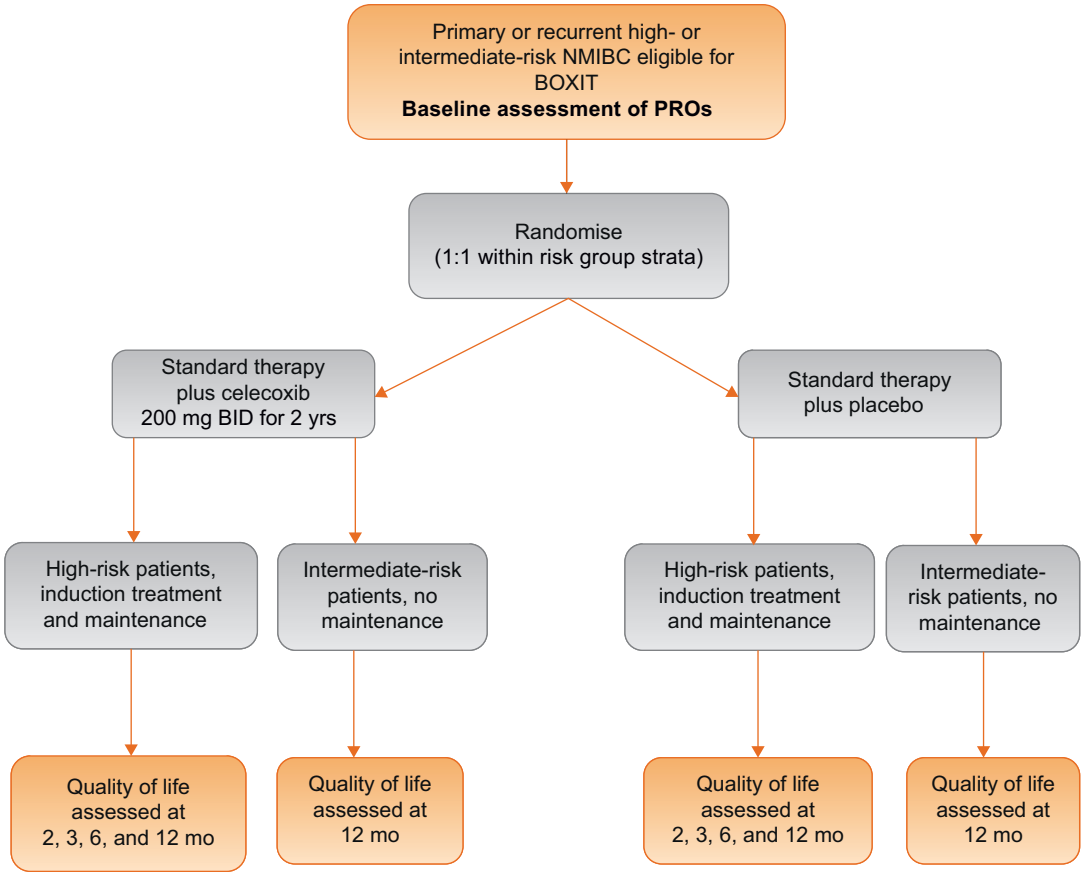

Fig. 1 – Trial schema with timing of assessments using the EORTC QLQ-C30 and QLQ-NMIBC24.
BOXIT = Bladder COX-2 Inhibition Trial; NMIBC = non–muscle-invasive bladder cancer; PRO = patient-reported outcome.

Response rates (based on entirely missing questionnaires or unusable questionnaire) at each time point were examined and reasons for missing questionnaires documented. Response rates to the sexual items were calculated based on whether patients reported being at least "a little" sexually active (item 48).

The module was developed according to standard EORTC Quality of Life Group guidelines, and translations followed standard procedures [4]. The module has 24 items originally hypothesised to form multi-item scales assessing urinary symptoms (items 1–7), intravesical treatment issues (items 10 and 11), future perspective (items 12–14), fever and feeling ill (items 8 and 9), and abdominal bloating and flatulence (BAF) (items 15 and 16), along with single items addressing different aspects of sexual functioning (items 17–24). All responses are linearly transformed from 0 to 100, with a high score indicating more symptoms or problems or better function for the functional scales. Ethics committee approval and written informed consent were obtained. The sample was determined by the patients within the BOXIT study up until November 2012.

2.2. Defining the scales within the module

Multitrait scaling analyses with data from each of the time points examined whether the individual items may be grouped into the hypothesised scales. The items assessing sexual functioning included two items to be completed by all patients, two items for completion by men only, and one item for completion by women only; also, there were three items completed by men or women reporting to be sexually active in the past 4 wk. Given the conditional nature of many of these items, it was not possible to analyse them as one scale.

Statistical evidence of *item convergent validity* was defined as a correlation of $\geq$0.40 between an item and its own scale (corrected for overlap) [10]. *Item discriminant validity* was defined as a correlation of $<$0.40 between an item and other scales in the questionnaire. An item was considered to be a scaling success when the correlation between the item and its own scale was greater than its correlation with any other scale. For each scale, the ceiling and floor effects were examined. After finalising the scale structure, other tests were performed.

2.3. Evaluating the reliability and validity of the module

The internal consistency was assessed by the Cronbach α coefficient, with $>$0.70 considered acceptable for group comparisons being examined within each scale at each assessment point [11].

Known group comparisons evaluated whether the module was able to discriminate between subgroups of patients differing in clinical status [11]. Known groups used for this comparison were baseline differences in QLQ-C30 physical function scores, with $<$90 or $>$90 representing relatively high (better) or relatively low (worse) scores, respectively. It was hypothesised that the scale scores of the QLQ-BLS24 would be higher (show more problems) in patients with lower physical function. Additional exploratory known groups validity testing was performed comparing data from men versus women. The independent student *t* test was used to examine differences in mean scores. Effect sizes were

Table 1 – Clinical details and questionnaire response rates

Clinical details	All patients, *n* = 410	High risk, *n* = 306	Intermediate risk, *n* = 104
Age, yr, mean (SD)	66.7 (9.3)	66.6 (9.7)	66.8 (7.8)
Age, yr, range	35–91	35–91	35–87
Gender male, no. (%)	325 (79.3)	247 (80.7)	78 (75.0)
Tumour grade, no. (%)			
G1	19 (4.6)	3 (1.0)	16 (15.4)
G2	149 (36.3)	61 (19.9)	88 (84.6)
G3	209 (51.0)	209 (68.3)	0 (0.0)
Unknown	33 (8.1)	33 (10.7)	0 (0.0)
Tumour stage, no. (%)			
Ta	167 (40.7)	78 (25.5)	89 (85.6)
T1	167 (40.7)	152 (49.7)	15 (14.4)
Tis	45 (11.0)	45 (14.7)	0 (0.0)
Ta/Tis	17 (4.1)	17 (5.6)	0 (0.0)
T1/Tis	14 (3.4)	14 (4.6)	0 (0.0)
Smoking status, no. (%)			
Current	127 (31.0)	102 (33.3)	25 (24.0)
Previous	213 (52.0)	159 (52.0)	54 (51.9)
Never	60 (14.6)	36 (11.8)	24 (23.1)
Diabetes present, no. (%)	32 (7.8)	22 (7.2)	10 (9.6)
Questionnaire response rates, no. (%)			
Baseline	401 (97.8)	298 (97.4)	103 (99.0)
2 mo*	282 (92.2)	282 (92.2)	N/A
3 mo*	288 (94.1)	288 (94.1)	N/A
6 mo*	263 (85.9)	263 (85.9)	N/A
12 mo	298 (86.1)	217 (94.3)	81 (77.9)
Response rate to sexual scales/items, no. (%)**			
Sexual function	1424 (93.0)	1248 (92.6)	176 (95.7)
Male sexual problems	1055 (85.8)	930 (85.1)	125 (91.9)
Sexual intimacy	505 (76.6)	445 (77.0)	60 (74.1)
Risk of contamination	504 (76.5)	444 (76.8)	60 (74.1)
Sexual enjoyment	498 (75.6)	439 (76.0)	59 (72.8)
Female sexual problems	70 (79.5)	57 (78.1)	13 (86.7)

N/A = not available; SD = standard deviation.
* Denominator for 2-, 3-, and 6-mo time points is 306 (high-risk patients only).
** Response rates for patients who are sexually active at each time point.

expressed as the mean difference divided by the pooled standard deviation (SD). Effect sizes were interpreted using the Cohen rule of thumb that a change of 0.5 SD represents a moderate effect, and a change >0.8 SD is a large effect [12].

To assess validity, correlations between the scales of the QLQ-BLS24 module and the scales of the QLQ-C30 were made using baseline data. Polychoric correlations were calculated, as is appropriate for items with four response categories. The responsiveness of the module to changes in health over time was examined in high-risk patients who underwent intensive treatments. It was hypothesised that during treatment, patients would report increased urinary symptoms and decreased generic aspects of health. Pairwise comparisons of changes in mean scores from baseline to 2, 3, 6, and 12 mo were evaluated using *t* tests for correlated samples. Because multiple comparisons were performed, a cautious but uncorrected *p* value of <0.01 was considered to be statistically significant.

All analyses were performed using Stata/IC statistical software (release 12, 2009; StataCorp LP, College Station, TX, USA).

3. Results

3.1. Patient characteristics, response rates, and missing data

At the time of data analyses, 472 patients were randomised, 433 patients consented to the quality-of-life study, and 410 of them completed a baseline questionnaire. Of these patients, 401 (97.8%) had complete baseline PRO data sets. The majority (79.3%) were men, and more than two-thirds had high-risk tumours (*n* = 306, 74.6%) (Table 1). The number of questionnaires returned at each time point and completion rates were 282 (92.2%), 288 (94.1%), 263 (85.9%), and 298 (94.3%) at 2, 3, 6, and 12 mo, respectively, for the high-risk group; at 12 mo, 81 questionnaires (77.9%) were returned for the intermediate-risk group. There were therefore 1532 questionnaires in total, with a completion rate for the five assessment points of 88.2%. At baseline, 48% of patients reported at least a little sexual activity (item 48), meaning that completion rates for the sexual scales and items were generally good (>75%). Sociodemographic and clinical details and questionnaire response rates are shown in Table 1.

3.2. Defining the scales in the module

Final results of the multitrait scaling analyses are shown in Table 2. Item within scale correlations in the original hypothesised urinary symptom, fever and malaise, and sexual function scales were all ≥0.40, and therefore these scales were maintained. Items 40 and 41, addressing intravesical treatment issues, showed many scaling errors. Discussion within the trial management group therefore led to agreement to include item 40 as a single item assessing intravesical treatment issues. Item 41 correlated well with the future perspectives scale, and therefore this scale was expanded to a four-item future worries scale (items 41–44). The two items assessing abdominal BAF (items 45 and 46) demonstrated satisfactory scaling properties when combined and thus formed a scale. The scale concerning sexual problems in men, items 49 and 50, functioned well and was retained. The remaining items in the original sexual function scale about sexual intimacy (item 51), risk of

Table 2 – Item convergent and discriminant correlations by scale within the EORTC QLQ-NMIBC at each follow-up time point*

Scale	Baseline assessment, *n* = 379 (all patients)				2-mo follow-up, *n* = 268*				3-mo follow-up, *n* = 260*				6-mo follow-up, *n* = 239*				12-mo follow-up, *n* = 270 (all patients)			
	Con	Dis	Test	α	Con	Dis	Test	α	Con	Dis	Test	α	Con	Dis	Test	α	Con	Dis	Test	α
US	0.50–0.77	−0.11 to 0.61	100	0.85	0.50–0.73	−0.13 to 0.48	100	0.89	0.46–0.77	−0.15 to 0.51	100	0.87	0.53–0.80	−0.17 to 0.53	100	0.88	0.53–0.81	−0.19 to 0.61	100	0.89
MAL	0.82–0.82	−0.25 to 0.43	100	0.57	0.82–0.82	−0.25 to 0.47	100	0.76	0.74–0.74	−0.10 to 0.53	100	0.58	0.79–0.79	−0.05 to 0.46	100	0.64	0.83–0.83	0.03–0.46	100	0.65
FW	0.70–0.85	0.16–0.33	100	0.90	0.69–0.85	−0.05 to 0.37	100	0.88	0.71–0.82	0.00–0.58	100	0.88	0.76–0.86	−0.09 to 0.49	100	0.89	0.77–0.87	0.10–0.44	100	0.91
BAF	0.58–0.58	−0.06 to 0.49	100	0.57	0.49–0.49	−0.18 to 0.26	90	0.56	0.61–0.61	−0.08 to 0.50	100	0.62	0.46–0.46	−0.13 to 0.46	100	0.49	0.58–0.58	0.59–0.00	90	0.58
SX	0.81–0.81	−0.16 to 0.10	100	0.82	0.82–0.82	–	100	0.83	0.84–0.84	−0.18 to 0.17	100	0.84	0.86–0.86	−0.15 to 0.15	100	0.86	0.89–0.89	−0.16 to 0.20	100	0.87
SXmen	0.76–0.76	−0.28 to 0.31	100	0.73	0.68–0.68	−0.39 to 0.22	100	0.71	0.74–0.74	−0.31 to 0.16	100	0.74	0.70–0.70	−0.31 to 0.27	100	0.70	0.75–0.75	−0.38 to 0.34	100	0.77

α = Cronbach α coefficient; BAF = bloating and flatulence; Con = the range of item-scale correlations (corrected for overlap); Dis = the range of correlations between an item and other scales; FW = future worries; MAL = malaise; SX = sexual function; SXmen = sexual problems in men; Test = the percentage of cases in which an item correlates equally or higher with its own scale than with other scales; US = urinary symptoms. NB responses to the SXmen scale were *n* = 288 at baseline, 202 at month 2, 196 at month 3, 177 at month 6, and 195 at month 12.

* At time points 2, 3, and 6 mo, only high-risk patients are included.

Table 3 – The scale structure of the EORTC QLQ-NMIBC24[†]

Originally hypothesised scales in the QLQ-BLS24	Items in each scale	Revised scales and single items in the QLQ-NMIBC24	Numbers of items in each scale/item
Urinary symptoms	31–37	Urinary symptoms	31–37
Malaise	38, 39	Malaise	38, 39
Intravesical treatment issues	40, 41	Intravesical treatment issues	40
Future worries	42–44	Future worries	41–44
Bloating and flatulence	45, 46	Bloating and flatulence	45, 46
Sexual function[*]	47–54	Sexual function[**]	47, 48
		Male sexual problems	49, 50
		Sexual intimacy	51
		Risk of contaminating a partner	52
		Sexual enjoyment[**]	53
		Female sexual problems	54

Scoring a high score is equivalent to more problems.
[†] Figure 2 shows the full questionnaire.
[*] Individual items.
[**] Scoring a high score is equivalent to better function.

contamination of partner (item 52), sexual enjoyment (item 53), and an item for sexual function in women only (item 54) remained as individual items.

At baseline, the revised scales showed some floor effects (as expected), because side-effects of treatment would be limited at that stage (scales for malaise, intravesical treatment >72% reported no problems at all). At all time points, few ceiling effects were noted (<2.5% for each scale; data not shown).

The original hypothesised scales in the EORTC QLQ-BLS24 and the confirmed scales in the EORTC QLQ-NMIBC24 are shown in Table 3.

3.3. *Reliability*

The internal consistency of the scales at each time point were good (>0.70) for the urinary symptoms, future worries, sexual function, and sexual function in men scales.

Table 4 – Mean patient-reported outcome scores in the QLQ-C30 and QLQ-NMIBC24 between patients with high and low performance status and between men and women

Scale/item	PF >90, *n* = 284, mean (SD)	PF <90, *n* = 110, mean (SD)	*p* value (*t* test)	Effect size[#]	Male, *n* = 316, mean (SD)	Female, *n* = 85, mean (SD)	*p* value (*t* test)	Effect size[#]
Functional scales, QLQ-C30[*]								
PF	98.8 (2.6)	77 (13.5)	<0.0001	2.94	93.3 (12.6)	90.5 (11.0)	0.066	0.23
Role function	96.5 (11.0)	77.7 (26.3)	<0.0001	1.12	90.9 (19.8)	92.2 (13.8)	0.588	−0.07
Emotional function	89.8 (13.7)	77.8 (21.0)	<0.0001	0.75	86.9 (17.1)	84.0 (16.6)	0.160	0.17
Cognitive function	92.1 (11.5)	82.3 (18.2)	<0.0001	0.72	89.4 (14.2)	89.3 (15.0)	0.962	0.01
Social function	92.6 (14.7)	77.5 (25.7)	<0.0001	0.81	87.6 (20.9)	92.0 (13.0)	0.066	−0.23
Global quality of life	83.5 (16.4)	67.3 (17.7)	<0.0001	0.98	79.5 (19.2)	77.9 (14.4)	0.498	0.08
Symptom scales, QLQ-C30[**]								
Pain	5.6 (11.7)	24.8 (26.2)	<0.0001	−1.13	11.0 (19.2)	10.6 (18.3)	0.858	0.02
Fatigue	7.9 (12.0)	27.4 (18.8)	<0.0001	−1.38	12.6 (16.9)	16.3 (15.6)	0.070	−0.22
Nausea and vomiting	0.6 (3.4)	3.9 (11.7)	<0.0001	−0.49	1.7 (7.9)	1.4 (4.6)	0.713	0.05
Module scales 24[**]								
Urinary symptoms	19.2 (17.0)	32.1 (21.1)	<0.0001	−0.71	23.8 (20.0)	19.6 (14.9)	0.072	0.22
Malaise	1.3 (5.3)	6.1 (13.0)	<0.0001	−0.59	2.6 (8.6)	2.6 (7.5)	0.949	0.01
Future worries	31.4 (23.0)	36.4 (26.2)	0.066	−0.21	33.0 (24.1)	32.3 (23.8)	0.830	0.03
Bloating and flatulence	14.0 (17.2)	17.7 (18.0)	0.055	−0.22	14.2 (17.0)	17.8 (18.7)	0.090	−0.21
Sexual function	27.3 (24.5)	13.7 (18.2)	<0.0001	0.60	26.5 (24.0)	11.9 (18.5)	<0.0001	0.64
Male sexual problems[a] (BL(BLSSXmen)	19.6 (27.6)	31.5 (36.2)	0.006	−0.40	22.5 (30.3)	NA	0.795	−0.17
Module single items[**]								
Intravesical treatment	8.5 (15.9)	13.1 (18.2)	0.013	−0.28	10.5 (17.3)	6.8 (13.5)	0.070	0.22
Sexual intimacy[b]	9.1 (19.4)	20.6 (35.8)	0.012	−0.49	10.8 (22.6)	14.1 (30.1)	0.518	−0.14
Risk of contamination[b]	19.1 (26.8)	17.8 (30.0)	0.814	0.05	20.2 (28.5)	13.0 (24.1)	0.254	0.26
Sexual enjoyment[b]	67.5 (30.1)	43.3 (32.9)	0.0002	0.79	65.4 (32.4)	49.3 (26.3)	0.025	0.51
Female sexual problems[c]	22.9 (26.4)	20.8 (35.4)	0.872	0.07	NA	NA	NA	NA

NA = not available; PF = physical function; SD = standard deviation.
[#] Effect size is mean difference divided by standard deviation.
[*] A higher score means better function.
[**] A high score means more symptoms or worse problems.
[a] Total number of respondents was 288 (91.1%).
[b] Total number of respondents was 128 (73%) answering questions about sexual intimacy, risk of contamination, and sexual enjoyment.
[c] Total number of respondents was 19 females (79%) answering questions about female sexual problems.

Table 5 – Validity–polychoric correlations between scales in the QLQ-C30 and the QLQ-NMIBC24

QLQ-C30 scales	Urinary symptoms	Malaise	Future worries	Bloating and flatulence	Sexual function	Sexual problems in men
Physical function	−0.29	−0.28	−0.07	−0.10	0.33	−0.22
Role function	−0.41	−0.61	−0.24	−0.22	0.14	−0.34
Emotional function	−0.25	−0.39	−0.50	−0.32	0.01	−0.08
Cognitive function	−0.29	−0.31	−0.16	−0.29	0.14	−0.24
Social function	−0.43	−0.52	−0.34	−0.15	0.20	−0.26
Global quality of life	−0.37	−0.46	−0.37	−0.21	−0.01	−0.04
Pain	0.44	0.47	0.18	0.33	−0.09	0.24
Fatigue	0.36	0.71	0.27	0.33	−0.18	0.24
Nausea and vomiting	0.26	0.59	0.15	0.35	−0.12	0.21

The fever and malaise scale had coefficients of >0.57 and 0.76; in the abdominal bloating scale, this ranged between 0.49 and 0.62 (Table 2).

3.4. Clinical validity

Patients with high scores (>90) on the physical function scale of the QLQ-C30 reported significantly better functional scores and fewer symptoms on all QLQ-C30 and three module scales (urinary symptoms, malaise, and sexual function) and on a single item (sexual enjoyment) than patients with poorer physical functioning ($p < 0.001$, Table 4). The effect sizes for these differences were moderate to large. Most scales and items were similar between men and women, except that men reported significantly more problems with sexual function ($p = 0.005$) than women (Table 4).

3.5. Criterion validity

The correlations between the majority of the scales in the core questionnaire and module ($n = 44$, 88%) were relatively low ($r < 0.40$, Table 5), indicating that the module is not overlapping in content with the QLQ-C30. Correlations >0.4 were observed between the malaise scale in the new module and role and social function scales, global quality of life, and the pain, fatigue, and nausea and vomiting scales in the QLQ-C30. The urinary symptoms scale was moderately associated with role

Table 6 – Responsiveness to change over time[†]

Function[*]	Baseline	2 mo	*p* value	3 mo	*p* value	6 mo	*p* value	12 mo	*p* value
Physical[*]	92.9	89.9	<0.001	90.3	<0.001	89.8	<0.001	89.7	<0.001
Role[*]	91.1	84.1	<0.001	86.8	<0.001	84.9	<0.001	87.2	0.008
Emotion[*]	86.7	84.9	0.097	85.0	0.107	86.8	0.877	87.2	0.757
Cognitive[*]	89.0	86.0	0.002	86.3	0.002	86.0	0.001	86.5	0.001
Social[*]	88.0	85.5	0.046	87.8	0.452	87.3	0.238	87.8	0.301
Global QOL[*]	78.5	75.1	0.003	75.7	0.016	74.2	0.003	74.9	0.001
Symptoms									
Fatigue	10.8	15.7	<0.001	19.2	0.033	14.7	0.007	13.3	0.039
N&V	13.7	21.3	<0.001	3.3	<0.001	20.2	<0.001	18.3	<0.001
Pain	1.7	3.0	0.040	13.8	<0.001	2.8	0.008	3.0	0.002
Dyspnoea	6.3	10.2	0.001	10.2	<0.001	10.5	<0.001	9.6	0.002
Sleep	18.0	20.4	0.115	19.2	0.341	22.1	0.006	20.7	0.004
Appetite	3.0	5.9	0.001	4.6	0.058	5.7	0.012	5.2	0.070
Cons	8.5	9.0	0.684	10.2	0.072	11.1	0.043	9.2	0.191
Diarrhoea	4.5	6.4	0.087	6.5	0.067	6.7	0.107	6.0	0.347
NMIBC24									
Urinary	23.4	26.2	0.040	22.8	0.4389	23.9	0.913	22.3	0.916
Malaise	3.1	9.3	<0.001	5.9	0.001	5.8	0.004	5.1	0.035
Future worries	33.3	30.0	0.011	29.3	0.002	28.2	0.001	26.1	<0.001
BAF	14.5	20.6	<0.001	18.2	0.001	20.0	<0.001	19.9	<0.001
SX	24.2	23.5	0.514	26.2	0.594	26.4	0.293	25.9	0.892
SXmen	22.4	28.1	0.016	24.2	0.147	25.4	0.149	28.8	0.006
Intravesical	10.1	12.5	0.094	10.2	0.739	10.7	1.000	9.6	0.416
SXI[**]	11.0	16.2	0.083	13.1	0.549	13.0	0.311	8.2	0.497
SXCP[**]	20.4	32.4	0.001	18.5	0.892	18.6	0.883	15.6	0.0132
SXEN[*,**]	70.7	64.0	0.707	67.5	0.236	67.1	0.083	69.9	0.311
SXfem[**]	26.7	30.0	0.591	33.3	0.594	48.1	0.0956	33.3	0.604

BAF = bloating and flatulence; Cons = constipation; N&V = nausea and vomiting; QOL = quality of life; SX = sexual function; SXCP = risk of contamination of partner; SXEN = sexual enjoyment; SXfem = sexual function in women; SXI = sexual intimacy; SXmen = sexual problems in men.
A high score means more problems except in function scales, in which a high score is equivalent to better function.
[*] Function scales, in which a high score is equivalent to better function.
[†] Mean QLQ-C30 and NMIBC24 scores before and after treatment in high-risk patients ($n = 260$).
[**] The number of responders varies according to subgroup and month; for example, month 2 versus baseline had 157 men and 10 females.

(0.41) and social function (0.43) and the pain scales (0.44) in the QLQ-C30, and the future worries scale in the module showed a moderate association with the emotional function scale (0.50).

3.6. Responsiveness to changes over time

Table 6 shows change in scores before and after treatment. Although little increase in urinary symptoms was observed during the follow-up period, several aspects of health measured by both the QLQ-C30 and the module did deteriorate during the first year of treatment. Significantly poorer physical, role, and cognitive function scores and worse nausea and vomiting and dyspnoea were seen at all time points. These findings were reflected in worse global quality-of-life scores at most assessments. Problems with malaise and abdominal bloating were observed at most follow-up assessments.

The final module was renamed the EORTC QLQ-NMIBC24 in keeping with current terminology (Fig. 2).

EORTC QLQ–NMIBC24

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

During the past week:	Not at all	A little	Quite a bit	Very much
31. Have you had to urinate frequently during the day?	1	2	3	4
32. Have you had to urinate frequently at night?	1	2	3	4
33. When you felt the urge to pass urine, did you have to hurry to get to the toilet?	1	2	3	4
34. Was it difficult for you to get enough sleep, because you needed to get up frequently at night to urinate?	1	2	3	4
35. Have you had difficulty going out of the house, because you needed to be close to a toilet?	1	2	3	4
36. Have you had any unintentional release (leakage) of urine?	1	2	3	4
37. Have you had pain or a burning feeling when urinating?	1	2	3	4
38. Did you have a fever?	1	2	3	4
39. Did you feel ill or unwell?	1	2	3	4
40. Did you have trouble arranging your life around the repeated bladder treatment appointments (cystoscopies or instillations)?	1	2	3	4
41. Did you worry about having repeated bladder treatments (cystoscopies or instillations)?	1	2	3	4
42. Were you worried about your health in the future?	1	2	3	4
43. Did you worry about the results of examinations and tests?	1	2	3	4
44. Did you worry about possible future treatments?	1	2	3	4
45. Did you have a bloated feeling in your abdomen?	1	2	3	4
46. Have you had flatulence or gas?	1	2	3	4
47. To what extent were you interested in sex?	1	2	3	4
48. To what extent were you sexually active (with or without sexual intercourse)?	1	2	3	4
49. For men only: Did you have difficulty gaining or maintaining an erection?	1	2	3	4
50. For men only: Did you have ejaculation problems (e.g. dry ejaculation)?	1	2	3	4
51. Have you felt uncomfortable about being sexually intimate?	1	2	3	4
52. Have you worried that you may contaminate your partner during sexual contact with the bladder treatment you have been receiving?	1	2	3	4
53. To what extent was sex enjoyable for you?	1	2	3	4
54. For Women only: did you have a dry vagina or other problems during intercourse?	1	2	3	4

Fig. 2 – The European Organization for Research and Treatment of Cancer module for non–muscle-invasive bladder cancer.

4. Discussion

This study evaluated the EORTC questionnaire module for NMIBC. An evidence-driven adaptation of the original scale structure into a revised module with six scales (urinary symptoms, malaise, future worries, bloating and flatulence, sexual function, and male sexual problems) and five single items was undertaken. Testing of the revised module yielded data supporting its clinical, construct, and criterion validity and acceptability of the module to patients (completion rates were high, with minimal missing data). The module was responsive to changes in health over time and was renamed the EORTC QLQ-NMIBC24 to reflect current terminology.

The purpose of measuring PROs in clinical trials alongside standard end points is to generate information to inform patients and their physicians about how treatments affect quality of life [13,14]. This information can supplement clinical outcome data in decision making. While some studies have examined PROs of treatment for BCa, there is a lack of data using condition-specific questionnaire modules [15,16]. Condition-specific measures are available for many cancer sites, and this module will add to the portfolio [3–6].

Although this was a large prospective study, it does have its limitations; primarily, it was performed within a single clinical trial and country. This study used clinical evidence to drive and make small modifications to the scale structure of the questionnaire. Further work examining the additional measurement properties of the questionnaire in other settings is still needed, including assessments of test–retest reliability and other clinical validation (eg, whether the module distinguishes between NMIBC and muscle-invasive disease). It is also necessary to examine the measurement properties of the module in patients with low-risk NMIBC.

There were very few problems with missing questionnaires, indicating that the module is acceptable for patients in a clinical trial. There were, however, more missing data for the items addressing sexual function. Health-related quality-of-life issues related to sexual function are assessed in a number of EORTC modules, and work is ongoing to develop a unified and comprehensive approach to assessing sexual issues in trials in oncology.

This study used an evidence-driven approach to adapt the scale structure of the EORTC module for NMIBC and explored its psychometric properties in a cohort of UK patients. Further testing in an international setting is still needed.

5. Conclusions

The revised module has well-defined scales and items, is acceptable for patients, and has encouraging psychometric properties. The questionnaires may be obtained by contacting the EORTC Quality of Life Department [4]. It is recommended that the module be used as a supplement to the QLQ-C30 in clinical trials to assess PROs in patients with NMIBC.

Author contributions: Jane M. Blazeby had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Blazeby, Fayers, Hall.
Acquisition of data: Kelly, Hall, Lloyd, Waters.
Analysis and interpretation of data: Blazeby, Fayers, Hall, Aaronson, Kelly.
Drafting of the manuscript: Blazeby, Fayers, Hall, Aaronson, Kelly.
Critical revision of the manuscript for important intellectual content: Blazeby, Fayers, Aaronson.
Statistical analysis: Fayers.
Obtaining funding: Hall, Blazeby, Kelly.
Administrative, technical, or material support: Blazeby.
Supervision: Blazeby, Fayers.
Other (specify): None.

Financial disclosures: Jane M. Blazeby certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Jane M. Blazeby received funding from the MRC ConDuCT Hub for trials methodology research. John D. Kelly received funding from the UCLH Biomedical Research Centre. John D. Kelly is the chief investigator and Emma Hall is the senior statistician for the investigator-initiated BOXIT sponsored by the Institute of Cancer Research and funded by Cancer Research UK (CRUK/07/004; C8262/A5669; C1491/A9895) and educational grants from Kyowa Hakko UK and Cambridge Laboratories.

Funding/Support and role of the sponsor: Trial recruitment was facilitated within centres by the National Institute for Health Research Cancer Research Network. Pfizer provided study medication free of charge within the BOXIT trial.

References

[1] Bladder cancer incidence statistics. Cancer Research UK Web site. http://info.cancerresearchuk.org/cancerstats/types/bladder/incidence/. Accessed September 2013.
[2] US Department of Health and Human Services, Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed September 2013.
[3] Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality of life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365–76.
[4] Manuals. European Organization for Research and Treatment of Cancer Web site. http://groups.eortc.be/qol/manuals. Accessed September 2013.
[5] Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993;11:570–9.
[6] Home page. Functional Assessment of Chronic Illness Therapy Web site. http://www.facit.org. Accessed September 2013.
[7] de Velde A, Fossa S, Hall R, Aaronson NK. Development of an EORTC module for patients with bladder cancer. EORTC Quality of Life Group internal report; 2004.
[8] BOXIT (Bladder COX-2 Inhibition Trial). http://www.controlled-trials.com/ISRCTN84681538. Accessed September 2013.
[9] Oosterlinck W, Lobel B, Jakse G, Malmstrom PU, Stockle M, Sternberg C. Guidelines on bladder cancer. Eur Urol 2002;41:105–12.

[10] Nunnally JC. Psychometric theory. New York, NY: McGraw-Hill; 1978.
[11] Fayers PM, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. Chichester, UK: Wiley; 2007.
[12] Cohen J. Statistical power analysis for the behavioural sciences, ed. 2. Hillsdale, NJ: Lawrence Erlbaum; 1988.
[13] Botteman MF, Pashos CL, Hauser RS, Laskin BL, Redaelli A. Quality of life aspects of bladder cancer: a review of the literature. Qual Life Res 2003;12:675–88.
[14] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013;309:814–22.
[15] Porter MP, Penson DF. Health related quality of life after radical cystectomy and urinary diversion for bladder cancer: a systematic review and critical analysis of the literature. J Urol 2005;173:1318–22.
[16] Sabichi AL, Lee J, Grossman B, et al. A randomised controlled trial of celecoxib to prevent recurrence of non-muscle invasive bladder cancer. Cancer Prev Res 2011;4:1580–9.