EUROPEAN JOURNAL OF CANCER 46 (2010) 1033–1040

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

An international validation study of the EORTC brain cancer module (EORTC QLQ-BN20) for assessing health-related quality of life and symptoms in brain cancer patients

Martin J.B. Taphoorn [a,], Lily Claassens [b], Neil K. Aaronson [c], Corneel Coens [b], Murielle Mauer [b], David Osoba [d], Roger Stupp [e], René O. Mirimanoff [e], Martin J. van den Bent [f], Andrew Bottomley [b], On behalf of the EORTC Quality of Life Group, and Brain Cancer, NCIC and Radiotherapy Groups*

[a] *Medical Centre Haaglanden, The Hague and VU University Medical Centre Amsterdam, The Netherlands*
[b] *EORTC, Brussels, Belgium*
[c] *The Netherlands Cancer Institute, Amsterdam, The Netherlands*
[d] *Quality of Life Consulting, West Vancouver, Canada*
[e] *Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland*
[f] *Erasmus University Medical Center, Rotterdam, The Netherlands*

ARTICLE INFO

Article history:
Received 7 August 2009
Received in revised form 24 December 2009
Accepted 14 January 2010
Available online 22 February 2010

Keywords:
HRQOL
Validation
Brain cancer
Psychometric analysis

ABSTRACT

Aims: The psychometric properties of the EORTC QLQ-BN20, a brain cancer-specific HRQOL questionnaire, have been previously determined in an English-speaking sample of patients. This study examined the validity and reliability of the questionnaire in a multi-national, multi-lingual study.

Methods: QLQ-BN20 data were selected from two completed phase III EORTC/NCIC clinical trials in brain cancer ($N = 891$), including 12 languages. Experimental treatments were surgery followed by radiotherapy (RT) and adjuvant PCV chemotherapy or surgery followed by concomitant RT plus temozolomide (TMZ) chemotherapy and adjuvant TMZ chemotherapy. Standard treatment consisted of surgery and postoperative RT alone. The psychometrics of the QLQ-BN20 were examined by means of multi-trait scaling analyses, reliability estimation, known groups validity testing, and responsiveness analysis.

Results: All QLQ-BN20 items correlated more strongly with their own scale ($r > 0.70$) than with other QLQ-BN20 scales. Internal consistency reliability coefficients were high (all $\alpha \geqslant 0.70$). Known-groups comparisons yielded positive results, with the QLQ-BN20 distinguishing between patients with differing levels of performance status and mental functioning. Responsiveness of the questionnaire to changes over time was acceptable.

Conclusion: The QLQ-BN20 demonstrates adequate psychometric properties and can be recommended for use in conjunction with the QLQ-C30 in assessing the HRQOL of brain cancer patients in international studies.

* *Corresponding author:* Address: Department of Neurology/Neuro-Oncology, Medical Centre Haaglanden, P.O. Box 432, Lijnbaan 32, 2501 CK, The Hague, The Netherlands. Tel.: +31 70 3302000; fax: +31 70 3303113.
E-mail address: m.taphoorn@mchaaglanden.nl (M.J.B. Taphoorn).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2010.01.012

1. Introduction

The incidence of brain tumours is low compared to other common cancer sites such as breast and lung. Primary brain tumours account for only 2–3% of all cancers in adults.[1] However, despite its low incidence, brain cancer is a disease characterised by severe symptoms and poor prognosis. Currently, efforts at developing new therapeutic strategies in brain cancer are focused on prolonging survival. However, these treatments may increase neurotoxicity[2] and thereby negatively affect health-related quality of life (HRQOL). Special concerns exist with regard to cognitive dysfunction and personality changes.[3] Consequently, clinical benefits of any new treatment should be weighed against treatment side-effects that adversely influence patients' HRQOL.

In assessing the HRQOL of brain cancer patients it is important to use well-validated instruments. Today, the most widely used brain cancer-specific HRQOL questionnaires are the brain subscale of the Functional Assessment of Cancer Therapy (FACT-Br)[4] and the EORTC brain cancer-specific Quality of Life Questionnaire (QLQ-BN20).[5]

The QLQ-BN20 has previously been tested in English-speaking samples, and has been demonstrated to have adequate psychometric properties.[5] The purpose of the study reported here was to examine the validity and reliability of the QLQ-BN20 in an international setting, in multiple languages.

2. Patients and methods

2.1. Patients

In this study we used data derived from two phase III EORTC/NCIC cancer clinical trials (EORTC Protocol 26951[6] and 26891[7]). These studies were merged in order to create a large dataset ($N = 941$) in an international and cross-cultural setting.

Trial 26951 included 368 patients with newly diagnosed highly anaplastic oligodendroglioma (i.e. anaplastic oligodendroglioma or anaplastic oligo-astrocytoma), who were randomly assigned to either six cycles of adjuvant Procarbazine-CCNU-Vincristine (PCV) chemotherapy beginning 4 weeks after postoperative external radiation therapy (RT) (experimental arm) or to postoperative RT only (control arm). Trial 26981 included 573 patients with newly diagnosed glioblastoma multi-forme (GBM), who were randomly assigned to either postoperative external RT in combination with temozolomide (TMZ) chemotherapy and subsequently six additional courses of TMZ (experimental arm) or postoperative RT only (control arm). In both studies, patients had been stratified by institution, performance status (WHO 0–1–2), age ($\leqslant$40 and >40 in trial 26951; <50 and $\geqslant$50 in trial 26981), and extent of resection at surgery (biopsy only versus debulking surgery/resection). Details of the two trials have been reported elsewhere,[6,7] as have the longitudinal HRQOL data.[8,9] Also, the prognostic value of HRQOL data has been published.[10,11]

2.2. Sample size

By merging the data from these two trials, an overall sample size of 941 brain cancer patients was obtained, of which 891 had at least one valid HRQOL measurement. This sample size exceeds the recommendation of at least 10 cases per item in multi-variate statistical modelling.[12]

2.3. Questionnaires and data collection

In both trials, HRQOL was assessed with the EORTC QLQ-C30 (version 3.0) and the EORTC QLQ-BN20. The QLQ-C30 is the EORTC's core HRQOL questionnaire that assesses a range of functional outcomes and symptoms relevant to a wide range of cancer populations.[13] The QLQ-BN20 was developed as a site-specific supplement to the QLQ-C30 for use amongst brain cancer patients undergoing chemotherapy or RT. It addresses symptoms that are specific to brain cancer or its treatment.[5]

The HRQOL assessments took place after surgery and before the start of RT (baseline), 4 weeks after the end of RT (after RT), and every 3 months after the initiation of chemotherapy until tumour progression (1st follow-up, 2nd follow-up, etc.). The three monthly follow-up HRQOL assessments in trial 26981 only applied to the first year following RT. A six monthly assessment interval was used thereafter.

2.4. Hypothesised scale structure

The QLQ-BN20 consists of four multi-item scales that address: future uncertainty (four items); visual disorder (three items); motor dysfunction (three items); and communication deficit (three items). Additionally, seven single items assess headaches, seizures, drowsiness, hair loss, itchy skin, weakness of legs, and bladder control. All items and scale scores of the QLQ-BN20 are linearly transformed to a 0–100 scale, with higher scores reflecting more severe symptoms.

Details of the development and initial field-testing of the QLQ-BN20 in an English-speaking population ($N = 105$) have been previously published.[5] The results generally supported the validity and reliability of the questionnaire.

2.5. Statistical analysis

2.5.1. Multi-trait scaling

Multi-trait scaling was employed to examine the hypothesised scale structure of the QLQ-BN20. To test for item-scale convergent validity, correlations of 0.40 or greater were sought between an individual item and its scale, corrected for overlap. Item-scale discriminant validity was examined by comparing the correlation of each item with its own scale versus other scales. An item was expected to correlate significantly higher (at least 2 standard errors) with its own scale than with other scales.

Additionally, the internal consistency reliability of the QLQ-BN20 multi-item scales was calculated with Cronbach's alpha coefficient. An $\alpha \geqslant 0.70$ is considered acceptable for group comparison.[14]

2.5.2. Clinical validity

The known groups validity of the QLQ-BN20 was tested by examining the extent to which the questionnaire could distinguish clearly between groups of patients. First, it was hypothesised that patients with higher PS (WHO 0/1) would report lower levels of physically-oriented symptoms than patients with lower PS (WHO 2 or 3). Second, patients with

higher MMSE[15] scores (⩾27) were expected to report fewer communication deficits than patients with lower MMSE scores (<27). Third, it was hypothesised that, because they were receiving additional, more intensive treatment, patients in the experimental treatment arms would report lower levels of future uncertainty than those in the standard treatment arms. Finally, patients with grade 1–3 CTC toxicity as rated by the physician for alopecia and rash were expected to report more problems with hair loss and itchy skin than patients without these toxicities. Group differences were assessed using the Wilcoxon rank sum test at multiple points in time.

To examine the responsiveness of the QLQ-BN20, changes in selected scales from baseline to first follow-up were examined in light of changes in performance status over the same period (stable/improved versus deteriorated) using Fisher's Exact test. Also, changes in future uncertainty scores were examined over time using ANOVA. It was expected that, as a result of successful coping, patients would report less uncertainty about the future over time.

Finally, the construct validity of the QLQ-BN20 was examined by calculating the correlations between its scales and those of the QLQ-C30. Although the QLQ-BN20 is intended to generate information not captured by the QLQ-C30, it was expected that a number of the scales from these questionnaires would be moderately correlated ($r > 0.40$) (e.g. the future uncertainty scale of the QLQ-BN20 with the emotional functioning scale of the QLQ-C30). Conversely, scales from the two questionnaires with less conceptual overlap (e.g. the motor dysfunction scale of the QLQ-BN20 with the cognitive functioning scale of the QLQ-C30) were expected to exhibit much lower correlations ($r < 0.40$). Correlations between these scales of these two questionnaires would be relatively modest.

3. Results

3.1. Patient characteristics

Of the total sample of 941 brain cancer patients from the two clinical trials, 891 (94.7%) completed at least one HRQOL assessment and were included in the analysis. These patients were enroled by 119 institutions from 15 countries. Baseline socio-demographic and clinical characteristics for these 891 patients (see Table 1) were very similar to those of the total sample (data not shown).

3.2. Questionnaire completion rates

Based on the total patient sample ($N = 941$) the baseline questionnaire completion rate was 83%. HRQOL completion rates for other time points ranged from 61% (after RT) to 66% (first follow-up). Missing QLQ-BN20 data were due, in part, to the absence of suitable translations into several languages (Hungarian and Finnish). The mean number of missing QLQ-BN20 items at baseline was 5.8%, with 'hair loss' (12.7%) and 'itchy skin' (8.5%) being the principal missing items.

3.3. QLQ-BN20 descriptive statistics

Mean scores and standard deviations at baseline and after RT for all QLQ-BN20 items are provided in Table 2. Items related to the future uncertainty scale had the highest mean scores at both time points. Both minimum (1) and maximum (4) response values were recorded at each time point in all items. However, most answers were skewed to the low end of the response scales as patients tended to report either 1 'not at all' or 2 'a little' for most symptoms or dysfunctions. At baseline and after RT, the 'seizures' item had the lowest mean score when compared to all other QLQ-BN20 items. Less than 15% of all patients reported seizures at all (i.e. during the week previous to each questionnaire completion). A clear score increase was observed with respect to 'hair loss' and 'itchy skin' after RT, in contrast to the other items that showed relatively stable mean scores over time.

3.4. QLQ-BN20 scale structure

Overall, the hypothesised QLQ-BN20 scale structure was supported by multi-trait scaling analyses. At both baseline and following RT, item-scale correlations indicated that each item correlated significantly stronger with its own scale (range = 0.71–0.92) than with other scales (range = 0.03–0.73) (Table 3). Inter-scale correlations indicated no problematic overlap between the QLQ-BN20 scales (data not shown). However, the 'weakness of legs' item and the motor dysfunction scale (three items) correlated moderately ($r = 0.57$ at baseline and 0.64 after RT). This was mainly attributable to 'weakness of legs' being correlated strongly to one item within the motor dysfunction scale: i.e. 'feeling unsteady on your feet' ($r = 0.68$ at baseline and 0.73 after RT, see Table 3). Feeling unsteady on your feet may be caused by weakness of one or both legs, as well as by weakness of one side of the body. An additional exploratory analysis was undertaken to control for corticosteroid use at trial entry, since the use of corticosteroids is related to experiencing weakness of both legs but not to (focal) motor dysfunction. However, use of corticosteroids had no impact on the correlation between either the QLQ-BN20 scale or item (data not shown). Finally, patients with a maximum score on the QLQ-BN20 item blurred vision reported to have no double vision at all in approximately one-third of all cases (10 out of 33), thereby diluting the overall visual disorder scale score. Nonetheless, the internal consistency reliability was acceptable ($\alpha \geqslant 0.71$).

The reliability (internal consistency) of the QLQ-BN20 met the preset criteria, with Cronbachs coefficient alpha ranging from 0.71 to 0.90 based on baseline data (Table 3). Similar results were observed at the post-RT assessment (data not shown).

3.5. Clinical validity

3.5.1. Known-group comparisons

Patients with a better WHO PS (score of 0 or 1) reported significantly lower scores (i.e. fewer symptoms and less dysfunction) than those with poorer PS (score of 2 or 3) on the majority of QLQ-BN20 scales/items, including future uncertainty, visual disorder, motor dysfunction, communication deficit, drowsiness, weakness of legs and bladder control (all p-values < 0.01) at both baseline and first follow-up (Table 4). No statistically significant differences between PS groups were observed for headaches, seizures, hair loss, or itchy skin.

Table 1 – Baseline socio-demographic and clinical characteristics of the sample N = 891 patients).

Variable	Treatment				Total (N = 891)
	RT/PCV (N = 175) N (%%)	95 RT (N = 175) N (%%)	98 RT (N = 269) N (%%)	RT+TMZ (N = 272) N (%%)	N (%%)
Age					
Median	48.5	49.8	55.9	55.7	52.9
Range	18.6–68.7	19.2–68.7	23.1–70.8	18.6–70.5	18.6–70.8
Sex					
1. Male	101 (57.7)	107 (61.1)	163 (60.6)	176 (64.7)	547 (61.4)
2. Female	74 (42.3)	68 (38.9)	106 (39.4)	96 (35.3)	344 (38.6)
PFS (WHO)					
0	73 (41.7)	56 (32.0)	104 (38.7)	109 (40.1)	342 (38.4)
1	75 (42.9)	89 (50.9)	132 (49.1)	127 (46.7)	423 (47.5)
2	26 (14.9)	28 (16.0)	33 (12.3)	36 (13.2)	123 (13.8)
Missing	1 (0.6)	2 (1.1)	0 (0.0)	0 (0.0)	3 (0.3)
Mini mental state examination					
Median	28.0	28.0	29.0	29.0	29.0
Range	14.0–30.0	6.0–30.0	7.0–30.0	2.0–30.0	2.0–30.0
Type of surgery					
Biopsy only	25 (14.3)	23 (13.1)	42 (15.6)	43 (15.8)	133 (14.9)
Partial resection	93 (53.1)	79 (45.1)	123 (45.7)	120 (44.1)	415 (46.6)
Total resection	57 (32.6)	73 (41.7)	104 (38.7)	109 (40.1)	343 (38.5)
Country					
The Netherlands	70 (40.0)	64 (36.6)	32 (11.9)	32 (11.8)	198 (22.2)
Canada	0 (0.0)	0 (0.0)	85 (31.6)	82 (30.1)	167 (18.7)
France	47 (26.9)	52 (29.7)	23 (8.6)	25 (9.2)	147 (16.5)
Italy	21 (12.0)	22 (12.6)	20 (7.4)	16 (5.9)	79 (8.9)
Germany	2 (1.1)	3 (1.7)	30 (11.2)	33 (12.1)	68 (7.6)
Belgium	5 (2.9)	5 (2.9)	21 (7.8)	22 (8.1)	53 (5.9)
Switzerland	0 (0.0)	0 (0.0)	23 (8.6)	23 (8.5)	46 (5.2)
United Kingdom	15 (8.6)	13 (7.4)	6 (2.2)	7 (2.6)	41 (4.6)
Austria	5 (2.9)	5 (2.9)	10 (3.7)	11 (4.0)	31 (3.5)
Spain	0 (0.0)	0 (0.0)	7 (2.6)	11 (4.0)	18 (2.0)
Israel	0 (0.0)	0 (0.0)	5 (1.9)	7 (2.6)	12 (1.3)
Hungary	3 (1.7)	7 (4.0)	0 (0.0)	0 (0.0)	10 (1.1)
Sweden	3 (1.7)	3 (1.7)	0 (0.0)	1 (0.4)	7 (0.8)
Finland	4 (2.3)	1 (0.6)	0 (0.0)	0 (0.0)	5 (0.6)
Australia	0 (0.0)	0 (0.0)	2 (0.7)	1 (0.4)	3 (0.3)
Poland	0 (0.0)	0 (0.0)	3 (1.1)	0 (0.0)	3 (0.3)
Slovenia	0 (0.0)	0 (0.0)	2 (0.7)	1 (0.4)	3 (0.3)

Mini Mental State Examination (MMSE) levels were significantly related to QLQ-BN20 communication deficit, visual disorder, motor dysfunction, weakness of legs and bladder control scores at baseline (all p-values < 0.01). Patients who scored below 27 on the MMSE reported greater difficulties on these scales than those who scored at least 27 ($p < 0.001$).

The hypothesis that patients would report lower future uncertainty due to receiving additional treatment (experimental arm) was not confirmed. Nor were CTC scores (grade 1–3) for alopecia or rash associated significantly with the QLQ-BN20 single item measures of hair loss or itchy skin after RT or first and second follow-up (data not shown).

3.5.2. *Responsiveness*

Changes in the physically-oriented QLQ-BN20 scale scores were examined in light of changes in PS scores. Patients whose PS deteriorated from baseline to first follow-up (scores shifting from 0–1 to 2–3) showed a significantly larger deterioration in motor dysfunction, drowsiness, visual disorder and bladder control than patients with a stable course of performance status over time (all $p < 0.005$). Significant differences were not found, however, for weakness of legs, seizures or headaches (Table 5A). As hypothesised, scores on the future uncertainty scale score improved (i.e. decreased) from baseline to first follow-up (on average, by 12 points; $p < 0.001$) (Table 5B).

3.5.3. *Relationship between the QLQ-BN20 and the QLQ-C30 core questionnaire*

The four QLQ-BN20 multi-item scales (future uncertainty, visual disorder, motor dysfunction and communication deficit), correlated relatively highly with certain QLQ-C30 functioning scales. The future uncertainty scale correlated quite strongly with the QLQ-C30 emotional functioning scale ($r = -0.645$) and, to a lesser degree, with the global quality of life and social functioning scales. The communication deficit scale correlated highly with the QLQ-C30 cognitive functioning scale ($r = -0.608$). The motor dysfunction scale correlated with the

Table 2 – Scales descriptive statistics.

QLQ-BN20 scales/single items	Number of forms	Mean score	SD	N (%) floor	N (%) ceiling	Normality
BFU (future uncertainty)	745	36.96	27.07	79 (10.1)	23 (3.0)	0.940
	464	*30.81*	*24.98*	*67 (13.9)*	*9 (1.9)*	*0.919*
BVD (visual disorder)	746	12.81	18.83	401 (51.5)	3 (0.4)	0.721
	462	*12.65*	*18.78*	*254 (52.8)*	*2 (0.4)*	*0.717*
BMD (motor dysfunction)	744	17.50	22.20	319 (40.9)	7 (0.9)	0.781
	459	*17.48*	*23.31*	*202 (42.0)*	*6 (1.2)*	*0.758*
BCD (communication deficit)	742	17.50	24.46	366 (47.0)	37 (4.7)	0.741
	460	*18.78*	*24.24*	*207 (43.0)*	*21 (4.4)*	*0.774*
BHA (headaches)	740	21.22	26.22	387 (49.7)	25 (3.2)	0.748
	457	*17.29*	*24.49*	*273 (56.8)*	*13 (2.7)*	*0.694*
BSE (seizures)	738	6.05	18.45	651 (83.6)	11 (1.4)	0.370
	455	*6.15*	*18.01*	*397 (82.5)*	*6 (1.2)*	*0.387*
BDR (drowsiness)	739	24.00	25.77	332 (42.6)	21 (2.7)	0.784
	459	*27.16*	*27.21*	*184 (38.3)*	*19 (4.0)*	*0.806*
BHL (hair loss)	680	9.36	22.57	555 (71.2)	21 (2.7)	0.470
	456	*29.31*	*34.60*	*225 (46.8)*	*51 (10.6)*	*0.777*
BIS (itchy skin)	713	9.16	20.53	570 (73.2)	11 (1.4)	0.503
	458	*16.59*	*25.26*	*290 (60.3)*	*14 (2.9)*	*0.674*
BWL (weakness of legs)	730	16.03	25.49	482 (61.9)	19 (2.4)	0.660
	458	*18.63*	*27.41*	*279 (58.0)*	*21 (4.4)*	*0.691*
BBC (bladder control)	740	8.38	20.93	616 (79.1)	14 (1.8)	0.456
	460	*9.93*	*23.61*	*373 (77.5)*	*17 (3.5)*	*0.476*

Baseline scores are reported in the first row and scores after RT in the second row, in *italics*.

Table 3 – Scale description, multi-trait scaling results and reliability (baseline *N* = 741; RT N = 462).

QLQ-BN20 scales/single items	Number of items	Item – own scale correlation	Item – other scale correlation	Cronbach's alpha
BFU (future uncertainty)	4	0.71–0.82	0.05–0.33	0.80
		0.74–0.84	*0.07–0.43*	*0.82*
BVD (visual disorder)	3	0.72–0.84	0.05–0.39	0.71
		0.71–0.86	*0.08–0.40*	*0.74*
BMD (motor dysfunction)	3	0.78–0.81	0.05–0.68	0.73
		0.80–0.85	*0.04–0.73*	*0.78*
BCD (communication deficit)	3	0.89–0.92	0.03–0.35	0.89
		0.90–0.92	*0.05–0.39*	*0.90*
BHA (headaches)	1	1.00	0.06–0.25	
		1.00	*0.01–0.37*	
BSE (seizures)	1	1.00	0.00–0.21	
		1.00	*0.05–0.31*	
BDR (drowsiness)	1	1.00	0.14–0.42	
		1.00	*0.08–0.38*	
BHL (hair loss)	1	1.00	0.04–0.32	
		1.00	*0.01–0.35*	
BIS (itchy skin)	1	1.00	0.00–0.32	
		1.00	*0.07–0.35*	
BWL (weakness of legs)	1	1.00	0.12–0.57	
		1.00	*0.11–0.43*	
BBC (bladder control)	1	1.00	0.02–0.31	
		1.00	*0.07–0.44*	

Baseline scores are reported in the first row and scores after RT in the second row, in *italics*.

Table 4 – Mean scores with standard deviation (SD) for baseline and 1st follow-up for clinically distinct groups.

QLQ-BN20 scales		Performance status[a]			Mini mental state examination[b]		
		High, N = 665, N = 332	Low, N = 112, N = 34	Wilcoxon	High, N = 525	Low, N = 229	Wilcoxon
BFU (future uncertainty)	Baseline	35.85 (26.77)	43.30 (28.10)	0.008	35.93 (26.95)	38.61 (27.57)	0.199
	1st follow-up	23.92 (21.01)	47.14 (27.24)	0.000			
BVD (visual disorder)	Baseline	11.89 (17.93)	18.09 (22.76)	0.008	11.02 (17.63)	17.15 (20.86)	0.000
	1st follow-up	10.19 (16.31)	24.73 (23.08)	0.000			
BMD (motor dysfunction)	Baseline	13.82 (18.38)	38.84 (29.49)	0.000	14.89 (21.30)	23.64 (23.11)	0.000
	1st follow-up	13.98 (20.24)	44.44 (32.18)	0.000			
BCD (communication deficit)	Baseline	15.50 (22.84)	28.75 (29.75)	0.000	12.49 (19.62)	29.01 (30.05)	0.000
	1st follow-up	17.53 (22.05)	41.32 (34.59)	0.000			
BHA (headaches)	Baseline	20.55 (25.86)	25.55 (28.07)	0.078	19.31 (24.46)	24.96 (28.32)	0.018
	1st follow-up	16.18 (23.32)	15.63 (25.38)	0.721			
BSE (seizures)	Baseline	5.88 (18.01)	7.17 (21.00)	0.643	5.04 (16.47)	8.45 (22.70)	0.085
	1st follow-up	5.08 (16.26)	5.75 (20.06)	0.954			
BDR (drowsiness)	Baseline	22.19 (24.61)	34.56 (29.72)	0.000	22.69 (24.67)	28.33 (28.15)	0.018
	1st follow-up	23.81 (25.40)	46.88 (36.77)	0.000			
BHL (hair loss)	Baseline	9.15 (21.85)	10.88 (26.83)	0.840	8.89 (22.04)	11.23 (24.43)	0.207
	1st follow-up	18.90 (29.01)	14.58 (23.85)	0.578			
BIS (itchy skin)	Baseline	8.22 (18.67)	15.15 (29.07)	0.072	9.05 (19.97)	10.08 (22.29)	0.794
	1st follow-up	12.43 (23.24)	19.79 (27.90)	0.090			
BWL (weakness of legs)	Baseline	13.68 (22.78)	30.13 (34.92)	0.000	14.35 (24.40)	20.22 (27.83)	0.006
	1st follow-up	14.16 (25.49)	38.54 (39.81)	0.000			
BBC (bladder control)	Baseline	6.98 (19.00)	16.67 (28.64)	0.000	6.89 (18.49)	12.25 (25.37)	0.003
	1st follow-up	5.84 (16.57)	30.21 (30.95)	0.000			

[a] WHO performance status. High: scores 0/1. Low: score 2 (baseline) or 2/3 (1st follow-up).
[b] High MMSE: score $\geqslant$27 and $\leqslant$30. Low MMSE: score <27.

Table 5A – Changes in mean scores over time of QLQ-BN20 physically-oriented scales by Performance Status evolution class.[a]

QLQ-BN20 scales	Change in scores by evolution of PS			
	Worsening N = 80 Mean (SD)[c]	Stable or improved N = 285 Mean (SD)	Total N = 365 Mean (SD)	Fisher's Exact test[b] p-Value
Headaches	7.18 (26.67)	−4.24 (28.79)	−1.71 (28.68)	0.0237
Motor dysfunction	9.06 (20.02)	−0.64 (21.62)	1.47 (21.63)	0.0047
Drowsiness	12.82 (24.79)	0.87 (27.44)	3.49 (27.29)	0.0025
Weakness of legs	6.99 (26.40)	0.00 (31.51)	1.50 (30.58)	0.1515
Visual disorder	3.42 (12.25)	−1.84 (20.87)	−0.69 (19.42)	0.0047
Bladder control	9.38 (26.21)	−1.16 (19.67)	1.13 (21.66)	<0.0005
Seizures	3.23 (23.92)	−1.59 (21.64)	−0.57 (22.19)	0.1586

[a] Evolution classes: 'worsening' (a shift from 0–1 to 2–3) or 'stable or improved' (remain in the same category or any shift upwards), over the period from baseline to first follow-up.
[b] QLQ-BN20 scale scores were categorised as 'worsening', 'stable' or 'improvement'.
[c] Any change with a positive value indicates a deterioration of the physical status, while any change with a negative value indicates an improvement of the physical status.

Table 5B – Change in QLQ-BN20 future uncertainty scale score from baseline to first follow-up.

QLQ-BN20 scale	Baseline N = 301 Mean (SD)	First follow-up N = 301 Mean (SD)	Total N = 602 Mean (SD)	Analysis of variance
Future uncertainty	37.35 (26.30)	24.97 (21.96)	31.16 (24.99)	<0.0001

QLQ-C30 physical functioning scale ($r = -0.556$), and to a lesser degree with role functioning scale. Finally, the visual disorder scale of the QLQ-BN20 correlated with the QLQ-C30 cognitive functioning scale ($r = -0.539$).

4. Discussion

The objective of this study was to generate detailed data on the psychometric properties of the EORTC QLQ-BN20 HRQOL questionnaire module when used in a diverse sample of patients with primary brain tumours participating in large scale, international clinical trials. Such large scale, international field-testing of the questionnaire is important, given that it, along with the EORTC core HRQOL questionnaire (the QLQ-C30), is one of the most widely used measures for assessing HRQOL in brain cancer patients.

Missing questionnaire data at the item level was a very modest problem, with the exception of the items on hair loss and itchy skin at baseline. This is probably due to the fact that such toxicity mainly occurs after the start of RT and, consequently, may be considered by the majority of patient as being not applicable at baseline. An increase in completion rates for the hair loss and itchy skin items during and following RT was observed. No significant correlation was found between QLQ-BN20 scores for hair loss and itchy skin, and CTC scores for alopecia and rash. This latter finding corresponds with earlier studies, suggesting there is not necessarily agreement between patient and clinician reports of symptoms or adverse effects.[16,17] Apparently, the alopecia and rash as observed by the doctor did not necessarily translate into complaints on these issues, perhaps reflecting their relative importance for brain tumour patients.

The distribution of QLQ-BN20 item/scale scores was skewed, with scores mainly at the lower end of the scale (indicating few symptoms or functional problems), which is related to the inclusion in both trials of patients with a relatively good performance status. In particular, the prevalence of seizures appeared to be low. Less than 15% of all patients reported this symptom at any time, which is not in accordance with other reported prevalence rates of seizures in brain tumour patients.[18,19] This may be due, in part, to the fact that the QLQ-BN20 asks about symptoms experienced during the past week.

The multi-trait scaling analysis confirmed the hypothesised scale structure of the QLQ-BN20 questionnaire and the reliability of the multiple item scales. However, the single 'weakness of legs' item correlated quite highly with the motor dysfunction scale (three items), and in particular with the item feeling 'unsteady on your feet'. Weakness of (both) legs is considered a side-effect of corticosteroids, while feeling unsteady refers to a balance problem, which may be caused by either myopathy due to corticosteroids or by motor dysfunction due to the disease itself. The analysis did not provide evidence for this conceptual difference in that the relatively high correlation could not be explained by those patients receiving corticosteroids at entry. Nevertheless, both symptoms have important clinical value and their current position in the QLQ-BN20 module should be maintained. Furthermore, the correlation between the items 'blurred vision' and 'double vision' in the visual disorder scale was lower than expected, which probably reflects the clear distinction made between these visual symptoms by patients.

The known-groups validity and the responsiveness of the QLQ-BN20 were generally supported by the data. Scores on at least four QLQ-BN20 scales varied significantly as a function of WHO performance status. In addition, significant differences in QLQ-BN20 communication deficit, visual disorder and motor dysfunction scale scores were observed as a function of MMSE levels. However, our hypothesis that patients in the experimental treatment arm would experience less future uncertainty levels than patients in the standard treatment could not be substantiated. This suggests that patients do not necessarily interpret experimental (and more intense) treatment as providing more certainty about the future.

Importantly, expected patterns of change were observed in QLQ-BN20 physically-oriented scales as a function of the WHO performance status course. This indicates that the questionnaire is responsive to changes in health status over time. Finally, as hypothesised, the QLQ-BN20 future uncertainty level decreased over time, which may reflect the effectiveness of coping strategies.

The QLQ-BN20 scales future uncertainty and communication deficit correlated relatively high with the QLQ-C30 scale emotional functioning and cognitive functioning, respectively. Indeed, from a conceptual viewpoint these relate closely, and high correlation rates are therefore not unexpected. At the same time, information on future uncertainty and communication deficit is complimentary.

The results of the current study are similar to those of the previous validation study of the QLQ-BN20 conducted amongst English-speaking patients only.[5] Internal consistency reliability estimates were about the same, and in both studies the QLQ-BN20 could distinguish clearly between patients differing as a function of performance status and mental status. Also, in both studies, deterioration over time in performance status was reflected in significant changes scores for at least four of the problem areas assessed by the QLQ-BN20.

In conclusion, this international, cross-cultural validation study confirmed that the current QLQ-BN20 module, available in over 15 languages, has adequate psychometric properties and is an appropriate tool for measuring HRQOL in both English-speaking and European primary brain tumour patients. Additional research is needed to confirm that the questionnaire is also appropriate for use amongst patients from other areas of the world, including (South) Asia and Africa. Therefore, we encourage its further use and refinements on different brain cancer populations in the future, such as patients with brain metastases.

Conflict of interest statement

Dr. Stupp has served on advisory boards for Bristol-Myers Squibb, Merck KGaA, Roche and Schering-Plough.

Dr. Mirimanoff has received speaking engagements including travel and accommodation from Schering-Plough.

Dr. Van den Bent has been a consultant for Schering-Plough, has acted as a member for their speakersbureau and has received research grants from Schering-Plough.

Acknowledgements

This publication was supported by Grants no. 2U10 CA11488-37 from the National Cancer Institute (Bethesda, Maryland, USA) and by a donation from the KWF Kankerbestrijding from the Netherlands through the EORTC Charitable Trust. This analysis is funded by the EORTC Quality of Life Group. Its content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Cancer Institute.

REFERENCES

1. US News & World Report: Brain Tumor Channel. Cleveland Clinic [cited 2007 July]. Available from: URL: <http://health.usnews.com/usnews/health/brain/brain_tumor/brain.about.htm>.
2. Taphoorn MJB, Bottomley A. Health-related quality of life and symptom research in glioblastoma multiforme patients. *Expert Rev Parmacoeconomics Outcomes Res* 2005;**5**:763–74.
3. Heimans JJ, Taphoorn MJB. Impact of brain tumour treatment on quality of life. *J Neurol* 2002;**249**:955–60.
4. Weitzner MA, Meyers CA, Gelke CK, Byrne KS, Cella DF, Levin VA. The functional assessment of cancer therapy (FACT) scale. Development of a brain subscale and revalidation of the general version (FACT-G) in patients with primary brain tumors. *Cancer* 1995;**75**:1151–61.
5. 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**:139–50.
6. Van den Bent MJ, Carpentier AF, Brandes AA, et al. Adjuvant procarbazine, lomustine, and vincristine improves progression-free survival but not overall survival in newly diagnosed anaplastic oligodendrogliomas and oligoastrocytomas: a randomized European Organisation for Research and Treatment of Cancer phase III trial. *J Clin Oncol* 2006;**24**:2715–22.
7. Stupp R, Mason WP, Van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* 2005;**352**:987–96.
8. Taphoorn MJB, Van den Bent MJ, Mauer MEL, et al. Health-related quality of life in patients treated for anaplastic oligodendroglioma with adjuvant chemotherapy: results of a European organisation for research and treatment of cancer randomized clinical trial. *J Clin Oncol* 2007;**25**:5723–30.
9. Taphoorn MJB, Stupp R, Coens C, et al. Health-related quality of life in patients with glioblastoma: a randomized controlled trial. *Lancet Oncol* 2005;**6**:937–44.
10. Mauer MEL, Taphoorn MJB, Bottomley A, et al. Prognostic value of health-related quality-of-life data in predicting survival in patients with anaplastic oligodendrogliomas from a phase III EORTC brain cancer group study. *J Clin Oncol* 2007;**25**:5731–7.
11. Mauer MEL, Stupp R, Taphoorn MJB, et al. The prognostic value of health-related quality-of-life data in predicting survival in glioblastoma cancer patients: results from an international randomized phase III EORTC Brain Tumour and Radiation Oncology Groups, and NCIC Clinical Trials Group study. *Br J Cancer* 2007;**97**:302–7.
12. Tabachnik BJ, Fidell LS. *Using multivariate statistics*. London: Harper and Row; 1993.
13. 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.
14. Fayers P, Machin D. *Quality of life: the assessment analysis and interpretation of patient reported outcomes*. Wiley: Chichester; 2007.
15. Folstein MF, Folstein SE, McHugh PR. Mini-mental state: a practical method for grading the cognitive state for the clinician. *J Psychiat Res* 1975;**12**:189–98.
16. Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J Clin Oncol* 2004;**22**:3485–90.
17. Basch E, Iasonos A, McDonough T, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol* 2006;**7**:903–9.
18. The Musella Foundation for Brain Tumor Research and Information. Clinical trials and Noteworthy treatments of Brain Tumors. Brain Tumor Symptoms [cited 2007 August]. Available from: URL: <http://virtualtrials.com/symptoms.cfm>.
19. Newton HB, Goldlust SA, Pearl D. Retrospective analysis of the efficacy and tolerability of levetiracetam in brain tumor patients. *J Neurooncol* 2006;**78**:99–102.