

Validity and reliability of the European Organization for Research and Treatment in Cancer Quality of Life Questionnaire (EORTC QLQ): experience from Kuwait using a sample of women with breast cancer

Shafika A. Alawadhi,^a Jude U. Ohaeri^b

From the ^aFaculty of Medicine, Kuwait University and the ^bDepartment of Psychiatry, Psychological Medicine Hospital, Safat, Kuwait

Correspondence: Jude U. Ohaeri, MD · Department of Psychiatry, Psychological Medicine Hospital, Gamal Abdul Naser Road, PO Box 4081, Safat, Kuwait 13041 · T/F: +965-2-489-9315 · judeohaeri@hotmail.com · Accepted for publication October 2009

Ann Saudi Med 2010; 30(5): 390-396

PMID: 20697165 DOI: 10.4103/0256-4947.67083

BACKGROUND AND OBJECTIVES: Although the EORTC QLQ-C30 and its breast-specific module (BR-23) are widely used instruments, the few reports on their psychometric characteristics from Arab and neighboring countries involved limited analyses. Our objective was to assess the psychometric characteristics of both questionnaires using the responses of a larger sample of Arab women.

METHODS: Participants were consecutive clinic attendees at the Kuwait Cancer Control Center. The indices assessed were alpha coefficients, item-internal consistency (IIC), item-discriminant validity (IDV), and known-groups validity.

RESULTS: The 348 women were aged 48.3 (10.3) years. The intra-class correlation for the test-retest statistic and the internal consistency values for the multi-item scales were >0.7 alpha. With the exception of the pain subscale, all items met the IIC criterion of >0.4 correlation with the corresponding scale. For IDV, the BR-23 performed better than the QLQ-C30. The scale scores discriminated between patients at different disease stages, and between sick and well populations.

CONCLUSION: With the exception of the pain subscale, the Arabic version of the questionnaires is psychometrically sound.

The effectiveness of modern methods of treatment and detection in oncology has contributed to the popular interest in quality of life (QOL) issues among survivors.¹ Although there has been considerable research on the QOL of women with breast cancer,² there is a paucity of such reports from the Arab and neighboring countries.³⁻⁶ Research interest in this field in the Arab world will be facilitated by the availability of psychometrically sound disease-specific instruments that have cross-cultural applicability. Towards this end, the European Organization for Research and Treatment in Cancer Quality of Life Questionnaire (EORTC QLQ-C30)⁷ and its breast-specific module (BR-23)⁸ are useful because they have been validated in diverse cultures,² and an official Arabic translation exists.

Previous reports on the validity and reliability of the EORTC QLQ-C30 and BR-23 from the Arab and neighboring countries^{3,5,6,9} have been based on relatively small sample sizes and the analyses were limited to a few issues of psychometrics. For instance, multi-trait scaling and test-retest reliability analyses were limited, and known-groups validity analysis did not involve comparison with a general population group.¹⁰

Our specific objectives were to assess (i) the test-retest reliability of the EORTC QLQ-C30 and BR-23; (ii) the floor/ceiling effect and acceptability of the items, as well as the internal consistency of the full questionnaires and their constituent multi-item scales; (iii) The item-internal consistency (IIC) and item-discriminant validity (IDV);¹¹ and (iv) the known-groups validity.¹⁰

METHODS

The subjects were consecutive attendees at the outpatient clinic of the medical oncology department of the Kuwait Cancer Control Center (KCCC) who fulfilled the study's inclusion criteria. They were attending follow-up clinic appointment for chemotherapy. Participants were in stable clinical condition and could independently give consent to participate in the study.⁷ In this culture, female patients are, as a rule, accompanied to hospital by family members who live with them.¹² Consent was also obtained from the family members, but the patients were interviewed privately in one of the clinic rooms. The KCCC is the national center for treatment of cancer. It has adequate facilities for diagnosis and treatment of cancer.

The EORTC QLQ-C30 is a 30-item generic health-related QOL instrument designed to assess cancer patients' physical, psychological and social functioning.^{7,13} It is composed of nine multi-item scales (5 functional scales, a global QOL scale [GQOL], and three symptom scales), five single-item symptom scales, and an item on the perceived financial impact of the disease (Table 1). In the version used for this study (version 3), the first 28 items are rated on a response scale of "not at all" (1), to "very much" (4).

The 23-item breast cancer-specific module, the QLQ-BR-23,⁸ consists of two multi-item functional scales, three multi-item symptom scales, and three single-item scales (Table 1). The response options are similar. The scoring algorithm recommended by the EORTC¹⁴ was used to transform the responses to values on a scale of 0% to 100%. For the functional scales and GQOL, a higher score corresponds to better functioning and QOL. For symptom scales, a higher score corresponds to more frequent and/or more intense symptoms.

The EORTC Quality of Life Unit in Belgium kindly sent us the questionnaires (English and Arabic translations). Ethical approval for the work was obtained from the institutional review panel of the KCCC. In addition, patients and their family caregivers gave verbal informed consent to participate in the study. They were duly informed that there would be no negative consequences for declining to participate. All families approached freely consented to participate in the study.

All assessments were based on private interviews by a trained female Arab research assistant. The criteria for staging disease by the doctors were those of the American Joint Committee on Cancer.¹⁵ Test-retest reliability was done by giving the questionnaires twice in a one-week period to 95 randomly selected literate, healthy, Kuwaiti women (aged >20 years, and married, to match the patients' socio-demographic characteristics).

Data were analyzed by SPSS, version 15 (SPSS Inc., Chicago, Illinois). The scale scores of the QLQ-C30 and BR-23 were computed as recommended.¹⁴ Data for test-retest reliability were analyzed by intra-class correlation coefficient (ICC), Kendall's tau correlation and kappa statistic for item agreements.¹⁶ The internal consistency was assessed by Cronbach's alpha. Acceptability of the questionnaires was assessed by the proportion of respondents who failed to complete each item. A cut-off value of <2.5% is recommended.¹¹ The proportion of respondents scoring at the lowest level (i.e., floor effect) and the highest level (ceiling effect) for each item was assessed. This is a measure of how far the item can be assumed to be capturing the full range of potential responses in the population.¹¹ Item internal consistency (IIC) and item discriminant validity (IDV), measured by Pearson's correlation, were assessed after adjusting for item overlap in the corresponding scale. The IIC and IDV concern the relationship of each item to its hypothesized scale or domain. The IIC rule requires that the item should correlate $r \geq 0.4$ with its adjusted scale score. For IDV, the item should have the highest correlation with its scale, in comparison with other scales in the questionnaire.¹¹

Known-groups validity was assessed, first by testing the significant differences in scale scores between subjects at different stages of the disease. Second, we compared scale score differences between the patients and the general population group, by effect size calculations and by adopting the operational definition of a clinically meaningful (significant) difference of 10% between groups.¹⁷ For effect size calculations, we defined a clinically significant difference as ≥ 0.5 .¹⁸ Missing data were automatically handled by the SPSS program by excluding cases analysis-by-analysis. The level of statistical significance was set at 5%.

RESULTS

In 2007 and 2008, 348 women fulfilled the inclusion criteria and agreed to participate. They were aged 20 to 81 years, with a mean and standard deviation of 48.3 (10.3) years. Six subjects (2.1% of 345) were aged <30 years, while 82 (23.8%) were aged >55 years. The majority (58.7%) were being treated for advanced disease (i.e., stages III and IV). The general population sample consisted of 95 women, aged 31(7.6) years (range: 23 to 55 years).

Reliability

The internal consistency values for the full questionnaires and their multi-item scales (i.e., ≥ 3 items) met the 0.7 Cronbach's alpha value requirement for the responses of the patients (Table 1). With regard to the

Table 1. Reliability analyses for QLQ C-30 and BR-23.

Variables/subscales	No. of subjects	No. of items	Cronbach's alpha	Floor effect (%) (range)	Ceiling effect (%) (range)
QLQ C 30 for first 28 items	253	28	0.94 ^a		
Physical functioning	330	5	0.79	3.3-14.4	7.5-20.9
Emotional	323	4	0.83	4.7-6.0	16.4-26.4
Fatigue	331	3	0.76	2.9-6.5	14.4-24.9
Body image	299	4	0.85	3.9-7.7	21.1-25.4
Systemic side effects	300	7	0.79	3.0-14.2	15.6-28.1
BR-23	171	23	0.92 ^b		
Breast symptoms	293	4	0.88	4.8-6.0	23.2-35.0
Arms symptoms	318	3	0.76	6.2-7.3	21.8-28.7
Role functioning	342, 344	2		5.2-5.3	10.9-12.9
Cognitive	343, 345	2		5.0-5.5	18.4-24.3
Social functioning	341	2		5.0-5.3	16.7-22.9
Nausea and vomiting	340	2		2.4-3.2	33.5-40.3
Pain	338, 343	2		5.2-5.7	13.6-15.2
Dyspnoea	335			6.0	18.8
Insomnia	338			7.1	18.3
Appetite loss	339			4.7	25.1
Constipation	345			3.5	42.0
Diarrhea	343			1.7	54.8
Financial difficulties	344			2.9	29.1
Global QOL	340			1.2,1.5	1.8
Sexual function	308, 326			0.6,1.5	31.3, 37.3
Sexual enjoyment	221			2.3	16.7
Future perspective	334			11.7	25.4
Upset by hair loss	331			10.9	17.2
Test-retest data ^c					
1st test:QLQ C-30 for first 28 items	77	28	0.93		
Test-retest for first 28 items	67	56	ICC:0.97 (0.95-0.98)		
2nd test: first 28 items of C-30	79	28	0.94		
Test-retest for general health and global QOL	92	4	ICC:0.93 (0.89-0.95)		
1st test BR-23	37	23	0.91		
Test-retest for BR-23	27	46	ICC: 0.94 (0.90-0.97)		

^aSplit-half reliability=0.84; ^bSplit-half reliability=0.81. ^cTest-retest data: Mean age=31(7.6), range 23-55 years, median=28, mode=24, N=95.

Table 2. Item-internal consistency and item-discriminant validity for QOLQ- C30.

Scale items	Physical (1)	Role (2)	Emotional (3)	Cognitive (4)	Social (5)
(1) Difficult effort	0.55	0.48	0.35	0.43	0.47
Problem walking long	0.61	0.49	0.29	0.35	0.39
Walking short	0.61	0.49	0.34	0.45	0.29
Need to stay in bed ^a	0.52	0.55	0.27	0.29	0.23
Need help eating	0.54	0.61	0.39	0.43	0.26
(2) Restrctied doing work ^b	0.59	0.43	0.41	0.43	0.37
Restricted hobbies ^b	0.57	0.43	0.41	0.41	0.33
(3) Nervous	0.32	0.39	0.65	0.55	0.47
Tension	0.32	0.38	0.69	0.65	0.49
Anxious ^a	0.37	0.39	0.61	0.67	0.45
Depressed	0.41	0.44	0.71	0.62	0.58
(4) Concentration difficulty ^a	0.48	0.47	0.66	0.49	0.48
Remembering difficulty ^a	0.40	0.39	0.70	0.49	0.53
(5) Family life ^a	0.34	0.38	0.55	0.51	0.49
Social life ^a	0.37	0.35	0.50	0.50	0.49
Symptoms scales	Fatigue (6)	Nausea and vomiting (7)	Pain (8)		
(6) Urgency to rest	0.61 ^d	0.38	0.44		
Exhausted	0.62	0.41	0.47		
Tired ^a	0.53	0.45	0.59		
(7) Have nausea	0.55	0.56	0.48		
Vomit	0.38	0.56	0.48		
(8) Feel pain ^c	0.49	0.40	0.39		
Miss activites b/c of pain ^c	0.43	0.63	0.39		

^aProbable scaling error: value of correlation <2SE with another scale, compared with own scale. ^bDefinite scaling error: value of correlation with another scale >2SE, compared with own scale. ^cDid not meet item internal consistency criterion. ^dFigures in bold belong to the same subscale.

floor and ceiling effects for the items of the QLQ-C30, the frequency of lowest scores was 0.6% to 14.4%, while the frequency of highest scores was 1.8% to 54.8%. For the QLQ-C30, 23 (76.7%) items had <2.5% missing values. For the BR-23, items from the following subscales had missing values: sexual (>6.3%), breast (>4.6%), body image (>3.2%), and upset by hair loss (9.2%).

The mean scores for the first 28 items of the QLQ-C30 (possible score: 1-4) ranged from 2.0 (0.75) to 2.73 (0.79) for 23 items, and for the two global QOL items (possible range: 1-7), it was 3.68-3.76. For the BR-23, the mean scores for 21 items (possible range: 1-4) was 2.06 (0.61) to 2.46 (0.95). The ICC for the

test-retest statistic (general population data) was highly significant for both the QLQ-C30 (0.97) and the BR-23 (0.94) (Table 1). Accordingly, the Kendall tau correlation coefficients for items in the test-retest data for the QLQ-C30 were high (all were >0.60, except five items with 0.5-0.59). In addition, the Kappa values indicated that all agreements were at least moderate (i.e., $k > 0.41$), with 10 being substantial (i.e., $k > 0.61$, $P < .0001$). A similar result was noted for the BR-23.

Item internal consistency and item discriminant validity
All the 15 items of the functional scales of the QLQ-C30 met the IIC requirement of correlation ≥ 0.4 (Table 2). With regard to IDV, there were definite

Table 3. Item internal consistency and item discriminant validity for QLQ-BR23.

Scale items	Body image (1)	Sexual functioning (2)	Systematic side effects(3)	Breast symptoms (4)	Arm symptoms (5)
(1) Feel unattractive	0.63	0.32	0.54	0.42	0.30
Feel unfeminine	0.71	0.19	0.49	0.37	0.25
Difficulty in looking at self	0.70	0.24	0.41	0.40	0.34
Unsatisfied with body	0.69	0.23	0.47	0.39	0.32
(2) Desire for sex	0.36	0.69	0.39	0.35	0.42
Sexually active	0.24	0.69	0.36	0.35	0.38
(3) Dryness of mouth	0.19	0.20	0.45	0.18	0.26
Taste food differently	0.23	0.23	0.47	0.33	0.27
Eye infection/pain	0.29	0.20	0.49	0.25	0.30
Hair loss	0.36	0.29	0.48	0.27	0.32
Sick/not well	0.47	0.22	0.59	0.37	0.33
Flush/red face ^a	0.52	0.29	0.49	0.42	0.42
Headache	0.55	0.26	0.62	0.39	0.39
(4) Pain affected breast	0.41	0.33	0.37	0.69	0.65
Breast swollen	0.40	0.30	0.44	0.72	0.65
Breast sensitive	0.32	0.32	0.36	0.72	0.56
Skin problems affected breast	0.45	0.33	0.45	0.78	0.59
(5) Pain in arm	0.27	0.35	0.36	0.49	0.55
Swelling in arm ^c	0.30	0.39	0.37	0.67	0.55
Difficulty moving arm	0.39	0.28	0.42	0.73	0.66 ^c

^aProbable scaling error: i.e., value of correlation <2SE with another scale, compared with own scale. ^bDefinite scaling error: value of correlation with another scale >2SE, compared with own scale. ^cFigures in bold belong to the same subscale.

scaling errors (see Table 2 for definition) for the following: (i) the two items of role functioning; and (ii) the two items of cognitive functioning. However, these scaling errors were conceptually logical because the items correlated with scales of similar construct (e.g., limitations in role functioning correlated with physical health). Similarly, in those items with probable scaling error (see Table 2 for definition), the correlations were conceptually logical.

Of the QLQ-C30 multi-item symptom scales, the two pain items just failed (0.39, each) to meet the IIC criterion, while one item (tiredness) had probable scaling error. The multi-item scales of the BR-23 fared better in IIC and IDV (Table 3). All the items met the IIC criterion of correlation >0.4 with their corresponding scale. There was only one definite scaling error (swelling in arm) and one probable scaling error (flush red

face), but these correlated higher only with items that were conceptually logical.

Known-groups validity

For the functional scales of the QLQ-C30, the general population women had significantly higher scores (i.e., better functioning) than the cancer patients, with a difference of at least 10% (effect size >0.5; except emotional scale: ES 0.33, 95%, CI=0.1-0.51). A similar pattern was evident for the multi-item symptom scales of the QLQ-C30, the body image functional scale of the BR-23 (ES 0.80, 95% CI=0.56-1.04), and the multi-item symptom scales of the BR-23 (ES >0.7). Using one-way ANOVA, we found that subjects with advanced disease tended to have worse functioning. This reached significance for the following: role functioning (stage IV <stage II, F=3.8, df=3/335, P<.01),

diarrhea (stage IV <stages I and II, $F=3.5$, $df=3/338$, $P=.02$), and future perspectives (stage III <stages I and II, $F=3.5$, $df=3/329$, $P=.02$).

DISCUSSION

The major limitation of the study is that the cross-sectional design did not allow us to test the sensitivity of the questionnaires to changes in clinical condition. However, our patients had similar demographic characteristics with those of breast cancer clinic populations in Kuwait.^{19,20}

The data on floor/ceiling effects and missing values indicate that the subjects responded to the full range of options and that the items were broadly acceptable and clear to them.⁸ The problem of reticence in responding to the sexual items was well noted in the original validation study of the BR-23 (consisting of Dutch, Spanish and United States samples), where it was stated that 11% to 14% of patients found one or more of these items to be too personal.⁸ This problem has also been noted in studies using other QOL instruments in the general population²¹ and clinical samples.²² Finally, it is to be noted that the item on sexual enjoyment is conditional on having been sexually active, while being upset by hair loss is conditional

on having experienced hair loss. Hence, missing values in these items are not good indices of the acceptability of the BR-23 questionnaire.⁸

The Arabic translation of both questionnaires generally met the statistical criteria for the reliability and validity issues investigated. With the exception of the two items of the pain subscale of the QLQ-C30, all the items of both questionnaires met the IIC criterion of >0.4 correlation with the corresponding scale. Coupled with the relatively low number of definite scaling errors, our data support the hypothesized subscales of the QLQ-C30 (except the pain subscale)¹³ and all the subscales of the BR-23.⁸

Our findings indicate that, with the exception of the pain subscale, the hypothesized subscales of the questionnaires are psychometrically sound in the Arab setting.

Acknowledgements

We thank Ms. Nashwa Al-Hassan for administering the questionnaires. Ms. Deepa Nair entered the data in SPSS. We thank Ms. Maha Alach for secretarial assistance and all the nurses and doctors at KCCC for their cooperation. We thank the patients and families for their cooperation.

REFERENCES

1. Penson RT, Wenzel LB, Vergote I, Cella D. Quality of life considerations in gynecologic cancer. *Int J Gynaecol Obstet* 2006;95:S247-57.
2. Montazeri A. Health-related quality of life in breast cancer patients: A bibliographic review of the literature from 1974 to 2007. *J Expt Clin Cancer Res* 2008;27:32.
3. Awad MA, Denic S, El Taji K. Validation of the European Organization for Research and Treatment in Cancer Quality of Life Questionnaire for Arabic-speaking populations. *Ann NY Acad Sci* 2008;1138:146-54.
4. Awadalla AW, Ohaeri JU, Gholoum A, Khalid AO, Hamad HM, Jacob A. Factors associated with quality of life of outpatients with breast cancer and gynecologic cancers and their family caregivers: A controlled study. *BMC Cancer* 2007;7:102.
5. Montazeri A, Harirchi I, Vahdani M, Khaleghi F, Jarvandi S, Ebrahimi M, et al. The EORTC breast cancer-specific quality of life questionnaire (EORTC QLQ-BR23): Translation and validation study of the Iranian version. *Qual Life Res* 2000;9:177-84.
6. Montazeri A, Harirchi I, Vahdani M, Khaleghi F, Jarvandi S, Ebrahimi M, et al. The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30): Translation and validation study of the Iranian version. *Support Care Cancer* 1999;7:400-6.
7. Kaasa S, Bjordal K, Aaronson N, Moum T, Wist E, Hagen S, et al. The EORTC core quality of life questionnaire (QLQ-C30): Validity and reliability when analysed with patients treated with palliative radiotherapy. *Eur J Cancer* 1995;31A:2260-3.
8. Sprangers MA, Groenvold M, Arraras JI, Franklin J, Te Velde A, Muller M, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: First results from a three-country field study. *J Clin Oncol* 1996;14:2756-68.
9. Guzelant A, Goksel T, Ozkok S, Tasbakan S, Aysan T, Bottomley A. The European Organization for Research and Treatment of Cancer QLQ-C30: An examination into the cultural validity and reliability of the Turkish version of the EORTC QLQ-C30. *Eur J Cancer Care (Engl)* 2004;13:135-44.
10. Hjermstad MJ, Fayers PM, Bjordal K, Kaasa S. Using reference data on quality of life—the importance of adjusting for age and gender, exemplified by the EORTC QLQ-C30 (+3). *Eur J Cancer* 1998;34:1381-9.
11. Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: The IQOLA project approach. *J Clin Epidemiol*. 1998;51:945-52.
12. Alshubaili AF, Awadalla AW, Ohaeri JU, Mabrouk AA. Relationship of depression and disability with the quality of persons with multiple sclerosis. *BMC Neurol* 2007;7:31.
13. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, 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 PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A. EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: European Organization for Research and Treatment of Cancer; 2001.
15. Greene FL, Page DL, Fleming ID. *AJCC cancer staging manual*. 6th ed. New York: Springer; 2002.
16. Viera AJ, Garrett JM. Understanding interobserver agreement: The Kappa statistic. *Family Med* 2005;37:360-3.
17. Sloane JA, Frost MH, Berzon R, Dueck A, Guyatt G, Moinpour C, et al. The clinical significance of quality of life assessments in oncology: A summary for clinicians. *Support Care Cancer* 2006;14:988-98.
18. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Med Care* 2003;41:582-92.
19. Saleh F, Abdeen S. Pathobiological features of breast tumours in the State of Kuwait: A comprehensive analysis. *J Carcinogenesis* 2007;6:12.
20. Abuzallouf S, Motawy M, Thotathil Z. Baseline staging of newly diagnosed breast cancer- Kuwait Cancer Control Center experience. *Med Prin Pract* 2007;16:22-4.
21. Ohaeri JU, Awadalla AW. The reliability and validity of the short version of the WHO Quality of Life Instrument (WHOQOL-Bref) in an Arab general population sample. *Ann Saudi Med* 2009;29:98-104.
22. Skevington SM, Lofty M, O'Connell KA. The World Health Organization's WHOQOL-Bref quality of life assessment: Psychometric properties and results of the international field trial. A report from the WHOQOL group. *Qual Life Res* 2004;13:299-310.