RESEARCH ARTICLE

Assessing the EORTC QLQ-BM22 Module Using Rasch Modeling and Confirmatory Factor Analysis across Countries: a Comprehensive Psychometric Evaluation in Patients with Bone Metastases

Chung-Ying Lin[1], Amir H Pakpour[2]*

Abstract

Background: The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Bone Metastases Module (EORTC QLQ-BM22) is a recently designed supplement to EORTC Quality of Life Questionnaire-C30 (EORTC QLQ-C30). Additional psychometric properties, especially using confirmatory factor analysis (CFA) and the Rasch model, are warranted. Materials and Methods: A total of 573 patients with bone metastases were enrolled from eight countries with a mean±SD age of 55.8±13.7 years. Slightly more than two thirds of them were female (n=383; 66.8%). CFA was used to examine the BM22 framework; Rasch models were applied to understand misfit items and differential item functioning (DIF). Results: The fit indices were satisfactory in CFA (comparative fit index=0.972, Tucker-Lewis index=0.964, root mean square error of approximation=0.076, and standardized root mean square residual=0.045). All items fit well in the Rasch models (mean square values were between 0.5 and 1.5), and only one item (number 17) displayed DIF across gender. However, there were six DIF items across Canada and Taiwan, ten across Canada and Iran, and six across Taiwan and Iran. Conclusions: The BM22 has satisfactory psychometric properties, and could assess the QoL of patients with bone metastases specifically focusing on their symptoms. Clinicians may want to use it to capture the underlying QoL for patients with bone metastases. However, the score of item 17 should be interpreted with caution when comparing male and female patients. In addition, researchers should note that variation in DIF items may occur when conducting an international study.

Keywords: Rasch analysis - confirmatory factor analysis - bone metastases - quality of life - EORTC QLQ-BM22

Introduction

Bone is the most common site for metastases, and bone metastases occur frequently in advanced cancer patients (Coleman, 2006), particularly in patients with breast, prostate, lung, and renal cell cancers, of whom 40% to 70% will eventually develop bone metastases (Coleman, 2006; Costa et al., 2008). Hence, clinicians should care for bone metastases in most cancer patients who frequently have bone pain or other significant complications such as spinal cord or nerve root compression, hypercalcemia or pathological fractures. As a result, their quality of life (QoL) is hugely jeopardized (Coleman, 2000; Costa et al., 2008).

QoL is one of the ultimately treatment endpoints for patients with bone metastases, as clinical trials reflect the importance of QoL (Chow and Bottomley, 2009; Zeng et al., 2012). Specifically, a recent survey updating

the international consensus on palliative radiotherapy endpoints for clinical trials in bone metastases stated that 91% of the interviewed radiation oncologists agreed the importance of incorporating a validated QoL instrument (Chow et al., 2012a). Furthermore, a general questionnaire for all cancer patients (e.g., the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-C30 [EORTC QLQ-C30]) lacks the ability to fully evaluate the specific aspects of QoL for patients with bone metastases (Chow and Bottomley, 2009). Therefore, the EORTC Bone Metastases Module (EORTC QLQ-BM22 or BM22) was designed as a supplement to the EQRTC QLQ-C30, and help clinicians to fully capture the QoL of a patient with bone metastases (Chow and Bottomley, 2009).

Although the BM22 was developed using sound methodology in accordance with the EORTC Quality of Life Group guidelines (Chow and Bottomley, 2009), we

*[1]Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, [2]Social Determinants of Health Research Center, Qazvin University of Medical Sciences, Qazvin, Iran *For correspondence: Pakpour_Amir@yahoo.com, apakpour@qums.ac.ir*

feel that psychometric properties are still insufficient for clinicians to use it. Specifically, the validation studies on BM22 (Chow et al., 2012b; Puskulluoglu et al., 2014; Yekaninejad et al., 2014) did not concern the wording effects (i.e., BM22 contains positively- and negatively-worded items) and primary used classical test theory (CTT). CTT is often criticized by its inappropriate treatment on scoring methods, such as means and standard deviations, and it does not focus on item parameters (i.e. item difficulty). Moreover, CTT relies on person parameters which depend on instrument items. Therefore, it is inappropriate to compare different instruments or populations using the CTT approach (Hobart and Cano, 2009).

Contrarily, Rasch model, a modern psychometric model belonged to the item-response theory (IRT) family, separately assesses person ability and item difficulty (DeRoos and Allen-Meares, 1993). Person ability refers to an individual's degree of a trait while item difficulty is the degree of a trait as determined by the item (DeRoos and Allen-Meares, 1993; Amin et al., 2012; Suhonen et al., 2013). Respondents with high person ability have a higher probability of success on a task than those with low person ability (DeRoos and Allen-Meares, 1993; Binda et al., 2013). In addition, Rasch can assess the validity for each individual item (i.e., testing the redundant and out-of-concept items), the threshold order of each item response, and differential item functioning (DIF) (Amin et al., 2012; Khan et al., 2013). The DIF items occurred when two groups (e.g., male and female) interpret the item differently. Therefore, it is important for clinicians to know which items display DIF. Also, researchers need to know which items display DIF across countries to make a valid international comparison.

The purpose of this study was to extend the knowledge of psychometric properties for BM22. In addition to some analyses using CTT, we used confirmatory factor analysis (CFA) to additionally account for the wording effect issue. We also applied Rasch models to evaluate the item difficulty, unfit items, disordered item responses, and DIF items for BM22.

Materials and Methods

Participants

We pooled and analyzed the data from two studies on psychometric evaluation of the EORTC QLQ-BM22 (Chow et al., 2012b; Yekaninejad et al., 2014). The participants recruitment procedure is detailed reported in the two studies (Chow et al., 2012b; Yekaninejad et al., 2014). In sum, our participants were those with bone metastases undergoing various treatments across eight countries (West countries: Canada, Brazil, France; East countries: India, Cyprus, Taiwan, Egypt, and Iran) with ten centers: Toronto, Ontario and Edmonton, Alberta (Canada); Kerala (India); Nicosia (Cyprus); Sao Paulo (Brazil); Taipei and Kaohsiung (Taiwan); Tanta (Egypt); Saint Etienne (France); Tehran (Iran). The recruitment periods were from March 2010 to January 2011 and from April 2012 to March 2013. Identical eligibility criteria for both recruitments were (1) patients > 18 years, (2)

histologically confirmed primary cancer, (3) radiological evidence of bone metastases, and (4) patients with the cognitively capability to complete the questionnaire in the respective languages (Chow et al., 2012b; Yekaninejad et al., 2014). All patients completed the EORTC QLQ-BM22 and a background information sheet (including sociodemographic and clinical characteristics) at baseline prior to treatment.
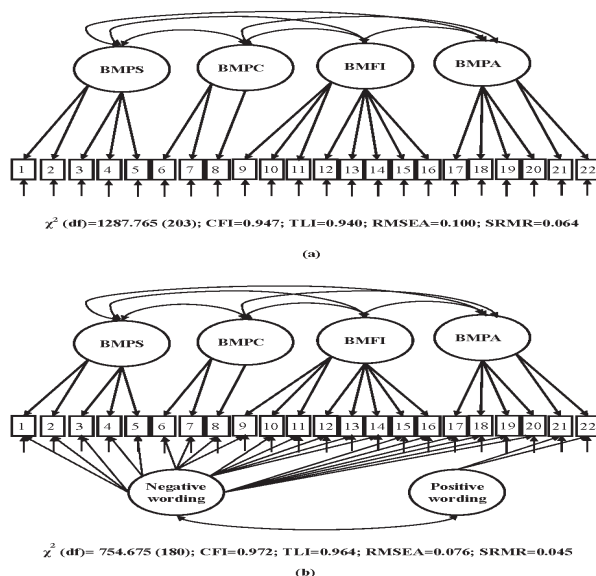
Questionnaire: EORTC QLQ-BM22

The EORTC QLQ-BM22 (BM22) is a 22-item questionnaire, developed as a supplement to EORTC QLQ-C30. All items are 4-point Likert scale from 1 (not at all) to 4 (very much), and the items are grouped into four domains: painful site (BMPS; 5 items), painful characteristic (BMPC; 3 items), functional interference (BMFI; 8 items) and psychosocial aspects (BMPA; 6 items). Scores of items 21 and 22 need to be reversed before calculating a 0-100 scale score: higher scores on BMPS, BMPC, and BMPA indicate poorer QoL; on BMFI indicate better functional ability (Fayers et al., 2001; Chow et al., 2012b).

The BM22 was translated into eight languages (Arabic, English, French, Greek, Malayalam, Mandarin for Taiwan, Farsi for Iran, and Portuguese for Brazil) of the participating countries based on EORTC translation guidelines. The processes included two individual forward translations, two individual backwards translations, and adequate pilot tests (Dewolf et al., 2009).

Statistical analysis

After reversely coding the items 21 and 22, we used the

χ^2 (df)=1287.765 (203); CFI=0.947; TLI=0.940; RMSEA=0.100; SRMR=0.064

(a)

χ^2 (df)= 754.675 (180); CFI=0.972; TLI=0.964; RMSEA=0.076; SRMR=0.045

(b)

Figure 1. (a) Confirmatory factor analysis (CFA) with four-correlated-trait framework; (b) CFA with four-correlated-trait and two-correlated-method framework. Abbreviations: CFI=comparative fit index; TLI=Tucker-Lewis index; RMSEA=root mean square error of approximation; SRMR=standard root mean square residual; BMPS=painful sites; BMPC=painful characteristics; BMFI=functional interference; BMPA=psychosocial aspects

BM22 raw scores to examine the ceiling and floor effects, internal consistency, confirmatory factor analysis (CFA), and Rasch models. The ceiling effects for domain and total scores of BM22 were reported using the percentages of the maximum response (i.e., the percentages of answering 4), floor effects using those of the minimum response. Based on Lin et al. (2013), we defined the ceiling and floor effects <3% as negligible, < 5% as minimal, and < 20% as moderate. The internal consistency was reported using Cronbachs' α, of which > 0.7 suggests satisfactory (Chang et al., 2014a).

In addition to using CFA to test the original framework of BM22 (see Figure 1a), we additionally tested another framework of BM22, which simultaneously accounting for trait and wording effects (see Figure 1b). Because BM22 contains two kinds of wording (two positively-worded items and twenty negatively-worded items), it is reasonable to test the data-model fit of a correlated-trait (i.e., four domains of BM22) correlated-method (i.e., positive and negative wordings) framework (Lin et al., 2014). Because all absolute skewness (0.071 to 1.257) < 3 and absolute kurtosis (0.154 to 1.321) <8 (Kline, 2005), we used the full information maximum likelihood estimation for the two CFA models. The fit indices for satisfactory data-model fit were nonsignificant χ^2, comparative fit index (CFI) and Tucker-Lewis index (TLI) > 0.9, root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) < 0.08 (Cheng et al., In press).

Rasch rating scale models were used to separately examine the unidimensionality of the four domains. Information-weighted fit statistic (infit) mean square (MnSq) and outlier-sensitive fit statistic (outfit) MnSq were used to determine whether an item is redundant to (infit or outfit MnSq < 0.5) or outside (infit or outfit MnSq > 1.5) the underlying concept (Lin et al., 2015). Moreover, the Rasch models provide separation reliability to estimate the reproducibility of the difficult hierarchy separately for the items and respondents. Unlike the traditional Cronbach's α, the person separation reliability is not affected by the item characteristics, and vice versa. Rasch models also provide separation index, which estimates the heterogeneity of the items' difficulty and respondents' ability. The recommended separation reliability is >0.7, and separation index is >2 (Chang et al., 2014b). In addition, we used average difficulty of each category,

step difficulty of each threshold or boundary between categories, and infit and outfit MnSq to examine the expected order. We anticipated monotonically increased average and step difficulties with infit and outfit MnSq between 0.5 and 1.5 (Jafari et al., 2012).

Rasch models tested the differential item functioning (DIF) for BM22 in terms of the following groups: gender (male vs. female) and countries (Canada, Taiwan, and Iran). The reason that we tested DIF for the three countries is because we need a sufficient sample size for each country, and a minimum of 200 is warranted based on a simulation study (Scott et al., 2009). However, some suggest that a sample size between 100 and 200 is somewhat acceptable (Lai et al., 2005). As a result, we decided using those countries with a sample size > 100 to detect DIF. In addition, the DIF contrast (the difficulty for Group 1 minus the difficulty for Group 2) >0.5 indicates a substantial DIF (Scott et al., 2009; Lin et al., 2015).

The CFA models were performed using LISREL 8.7 (Scientific Software International, Lincolnwood, IL, USA); Rasch using WINSTEPS (Linacre and Wright, 2009); all other analyses using SPSS 15.0 (SPSS Inc., Chicago, IL, USA).

Results

The eligible respondents were 117 in Canada, 11 in Egypt, 6 in France, 12 in Brazil, 19 in Cyprus, 224 in Taiwan, 7 in India, and 177 in Iran. Their mean±SD age was 55.81±13.69 years, and slightly more than two thirds of the participants were female (n=383; 66.8%).

The ceiling and floor effects for all domains of BM22 were negligible to moderate (ceiling effect: 0.5% to 3.5%; floor effect: 0.9% to 19.7%), and there were no ceiling and floor effects for the total score. Cronbach's α was acceptable for all domains and the total score (0.72 to 0.93), except for the BMPS domain (0.68). In addition, separation reliability was satisfactory for item separation (0.97 to 0.99) but not for person separation (0.55 to 0.85), and a similar trend was found in separation index (item separation=6.02 to 9.13; person separation=1.11 to 2.40) (Table 1).

The fit indices of CFA for the four-trait-correlated framework were all adequate, except for a slightly high value of RMSEA (0.100). The framework simultaneously accounting for the trait and method effects (i.e., the four-

Table 1. Internal Consistency, Separation Reliability and Separation Indices of BM22

| | Domain | | | | BM22 |
	BMPS	BMPC	BMFI	BMPA	
Item No.	5	3	8	6	22
Classical test theory					
Ceiling %	0.50%	2.10%	3.50%	1.00%	0.00%
Floor %	8.40%	19.70%	8.70%	0.90%	0.00%
Cronbach's α	0.68	0.77	0.93	0.72	0.93
Rasch					
Person separation index	1.11	1.38	2.4	1.39	--
Person separation reliability	0.55	0.66	0.85	0.66	--
Item separation index	6.02	7.93	5.54	9.13	--
Item separation reliability	0.97	0.98	0.97	0.99	--

BMPS=painful sites; BMPC=painful characteristics; BMFI=functional interference; BMPA=psychosocial aspects

Table 2. Confirmatory Factor Analysis (CFA) Standardized Factor Loadings and Rasch Difficulties and Rasch Fit Statistics for Each Item

Domain	CFA		Rasch		DIF contrast			
Item # and description	Loadings	Difficulties	Infit MnSq	Outfit MnSq	Male vs. Female	Canada vs. Taiwan	Canada vs. Iran	Taiwan vs. Iran
BMPS (Painful sites)								
1. Pain in your back	0.63	-0.56	0.93	0.93	0	-0.46	-0.70	-0.23
2. Pain in your leg(s) or hips(s)	0.62	-0.25	1.09	1.08	-0.18	-0.43	-0.20	0.22
3. Pain your arm(s) or shoulder(s)	0.46	0.13	0.99	1.01	0.09	0.58	0.75	0.17
4. Pain in your chest or rib(s)	0.55	0.12	0.96	0.95	0.26	0.56	0.16	-0.40
5. Pain in your buttock(s)	0.58	0.56	1.12	1.03	-0.15	-0.12	0.2	0.31
BMPC (Painful characteristics)								
6. Had constant pain	0.82	-0.17	0.94	0.92	-0.10	-0.09	0.46	0.55
7. Had intermittent pain	0.67	-0.69	1.08	1.06	0.4	0.29	-0.83	-1.12
8. Had pain not relieved by pain medications	0.66	0.87	1.02	0.92	-0.35	-0.30	0.36	0.66
BMFI (Functional interference)								
9. Had pain while lying down	0.76	0.44	1.05	1	0.17	0.43	0.11	-0.32
10. Had pain while sitting	0.77	0.56	0.79	0.83	0.1	0.1	-0.58	-0.67
11. Had pain when trying to stand up	0.78	0.26	0.85	0.82	-0.09	-0.03	-0.51	-0.48
12. Had pain while walking	0.81	0.2	0.84	0.84	0.06	0.17	0.02	-0.15
13. Had pain with activities such as bending or climbing stairs	0.8	-0.32	0.85	0.85	0.03	0.01	0.34	0.34
14. Had pain with strenuous activity (e.g., exercise, lifting)	0.71	-0.59	0.93	0.89	-0.09	0.04	0.47	0.43
15. Has pain interfered with sleeping at night	0.72	0.05	1.23	1.24	0.06	0.62	0.9	0.28
16. Had to modify your daily activities because of your illness	0.67	-0.60	1.43	1.48	0.26	-1.17	-0.67	0.51
BMPA (Psychosocial aspects)								
17. Felt isolated from those close to you (e.g., family, friends)	0.54	1.06	1.23	1.1	-0.53	-1.33	0.51	1.84
18. Worried about loss of mobility because of your illness	0.85	0	0.78	0.75	-0.18	-0.39	-0.38	0.01
19. Worried about becoming dependent on others because of your illness	0.84	-0.06	0.76	0.76	0	-0.36	-0.66	-0.29
20. Worried about your health in the future	0.68	-0.67	0.76	0.77	0.08	-0.07	-0.49	-0.42
21. Felt hopeful your pain will get better	0.29	-0.19	1.29	1.3	0.42	1.32	0.97	-0.34
22. Felt positive about your health	0.24	-0.15	1.27	1.33	0.1	0.24	0.02	-0.22

Infit MnSq=information-weighted fit statistic mean square; Outfit MnSq=outlier-sensitive fit statistic mean square; DIF=differential item functioning. DIF contrast=Difficulty of the former group (e.g., male) subtracts difficulty of the later group (e.g., female). Substantial DIF contrasts are in bold.

trait-correlated and two-method-correlated framework) had all fit indices satisfactory (Figure 1). In addition, the standardized factor loadings of all items were reported on Table 2.

All items had acceptable Rasch fit statistics in terms of infit and outfit MnSq, and their item difficulties fell between -0.56 to 0.56 for BMPS; -0.69 to 0.87 for BMPC; -0.60 to 0.56 for BMFI; -0.67 to 1.06 for BMPA (Table 2). If we additionally considered the step difficulties, the difficulty ranges were -2.36 to 2.28 for BMPS; -3.37 to 3.22 for BMPC; -3.15 to 3.01 for BMFI; -2.14 to 2.15 for BMPA. Moreover, all the step difficulties were monotonically increased by their categories with satisfactory infit and outfit MnSq (Table 3).

Only item 17 (Felt isolated from those close to you) displayed substantial DIF across gender (DIF contrast=-0.53). However, six substantial DIF items were found across Canada and Taiwan, ten across Canada and Iran, and six across Taiwan and Iran (Table 2). In addition, items 16 (Had to modify your daily activities because of your illness) and 17 displayed substantial DIF across the three countries.

Discussion

The satisfactory internal consistency, including Cronbach's α and item separation reliability, agrees with previous studies (Chow et al., 2012b; Yekaninejad et al., 2014). Therefore, the BM22 items are reliable for measuring QoL for patients with bone metastases. In addition, the low ceiling and floor effects, adequate fit indices in CFA models, good fit statistics in Rasch

models, and anticipated response ordering suggest that the psychometric properties of the BM22 are sound. Although the person separation index and reliability were low, and some DIF items were found, we feel that these are acceptable.

The low person separation index indicates that our sample was homogeneous in QoL, and low person separation reliability indicates unstable QoL performance (Chang et al., 2014b; Lin et al., 2015). However, the BM22 items had heterogeneous difficulties and stable item properties as indicated by our high item separation index and reliability. In addition, our results confirm the findings of Yekaninejad et al. (2014) that the framework of BM22 is valid. We additionally extend their findings that a slightly wording effect was found in the framework of BM22. A questionnaire containing both positively- and negatively-worded items may influence its psychometric properties (Lin et al., 2014) and our study found the impacts, though not substantial, in the BM22. Nevertheless, our results indicated that the construct of BM22 is supported.

An interesting finding is the DIF items in BM22. DIF occurs when the groups interpret the same item differently (Lin et al., 2015), and makes comparing DIF items across groups meaningless. Therefore, the clinicians should keep in mind that which items can be and which cannot be compared across certain groups. According to our results, all BM22 items but item 17 (Felt isolated from those close to you) were interpreted similarly across male and female patients with bone metastases. The DIF contrast was -0.53, which indicates that female patients tended to rate a higher score than did male patients on item 17 when both gender patients have the same isolated feeling. Some DIF items were found across Canada, Taiwan, and Iran. Nearly half of the BM22 items displayed DIF across Canada and Iran, while only six items were DIF across Canada and Taiwan, and across Taiwan and Iran. This may due to their cultures: Taiwan and Iran share some Asian values; Taiwan is a more westernization society than Iran is. As a result, Canadians and Iranians interpret most BM22 items differently. Another possibility is the lack of translation equivalence (Scott et al., 2010) though we are confident that it might impact little on the DIF because the translation process is robust and sound (Dewolf et al., 2009). In addition, we adopted the DIF contrast of 0.5, which represents an odds ratio of 1.65, as substantial DIF. Therefore, the DIF items displayed in our results could be attributable to both culture differences and, if there was, the lack of translation equivalence. Nevertheless, future studies are warranted to examine the impacts of culture differences in the DIF items after accounting for the translation equivalence.

The strength of this study is the international recruitment and the use of advanced psychometric methods. However, one major limitation of this study is no criterion was applied to test the concurrent validity of the BM22.

In conclusion, the BM22 is a sound instrument to assess the QoL for patients with bone metastases specifically on their symptoms. Clinicians could use it to well understand the underlying QoL in terms of painful sites, painful characteristics, functional interference, and psychosocial aspects for patients with bone metastases. However, they need to interpret the score of item 17 with cautions when comparing male and female patients.

References

Amin L, Rosenbaum P, Barr R, et al (2012). Rasch analysis of the PedsQL: an increased understanding of the properties of a rating scale. *J Clin Epidemiol*, **65**, 1117-23.

Binda D, Vanhoutte EK, Cavaletti G, et al (2013). Rasch-built Overall Disability Scale for patients with chemotherapy-induced peripheral neuropathy (CIPN-R-ODS). *Eur J Cancer*, **49**, 2910-8.

Chang CC, Wu TH, Chen CY, et al (2014a). Psychometric evaluation of the internalized stigma of mental illness scale for patients with mental illnesses: measurement invariance across time. *PLoS One*, **9**, 98767.

Chang KC, Wang JD, Tang HP, et al (2014b). Psychometric evaluation, using Rasch analysis, of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment: further item validation. *Health Qual Life Outcomes*, **12**, 148.

Cheng CP, Luh WM, Yang AL, et al (In press). Agreement of children and parents scores on Chinese version of Pediatric Quality of Life Inventory 4.0: further psychometric development. *Appl Res Qual Life*.

Chow E, Bottomley A (2009). Understanding the EORTC QLQ-BM22, the module for patients with bone metastases. *Expert Rev Pharmacoecon Outcomes Res*, **9**, 461-5.

Chow E, Hoskin P, Mitera G, et al (2012a). Update of the international consensus on palliative radiotherapy endpoints for future clinical trials in bone metastases. *Int J Radiat Oncol Biol Phys*, **82**, 1730-7.

Chow E, Nguyen J, Zhang L, et al (2012b). International field testing of the reliability and validity of the EORTC QLQ-BM22 module to assess health-related quality of life in patients with bone metastases. *Cancer*, **118**, 1457-65.

Coleman RE (2000). Management of bone metastases. *Oncologist*, **5**, 463-70.

Coleman RE (2006). Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin Cancer Res*, **12**, 6243-9.

Costa L, Badia X, Chow E, et al (2008). Impact of skeletal complications on patients' quality of life, mobility, and functional independence. *Support Care Cancer*, **16**, 879-89.

DeRoos YS, Allen-Meares P (1993). Rasch analysis. *J Soc Serv Res*. 161-7.

Dewolf L, Koller M, Velikova G, et al 2009. Translation procedure, brussels, belgium, EORTC Quality of Life Group.

Fayers P, Aaronson NK, Bjordal K, et al 2001. The EORTC QLQC30 Scoring Manual, Brussels, Belgium, European Organization for Research and Treatment of Cancer.

Hobart J, Cano S (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*, **13**, 1-177.

Jafari P, Bagheri Z, Safe M (2012). Item and response-category functioning of the Persian version of the KIDSCREEN-27: Rasch partial credit model. *Health Qual Life Outcomes*, **10**, 127.

Khan A, Chien CW, Brauer SG (2013). Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. *J Clin Epidemiol*, **66**, 681-7.

Kline RB 2005. Principles and practice of structural equation modeling, New York, Guilford Press.

Lai JS, Teresi J, Gershon R (2005). Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Eval Health Prof*, **28**, 283-94.

Lin CY, Luh WM, Cheng CP, et al (2014). Evaluating the

Chung-Ying Lin and Amir H Pakpour

wording effect and psychometric properties of the Kid-KINDL: using the multitrait-multimethod approach. *Eur J Psychol Assess*, **30**, 100-9.

Lin CY, Luh WM, Cheng CP, et al (2013). Measurement equivalence across child self-reports and parent-proxy reports in the Chinese version of the pediatric quality of life inventory version 4.0. *Child Psychiatry Hum Dev*, **44**, 583-90.

Lin CY, Yang SC, Lai WW, et al (2015). Rasch models suggested the satisfactory psychometric properties of the World Health Organization Quality of Life-Brief among lung cancer patients. *J Health Psychol*.

Linacre JM, Wright BD 2009. A user's guide to WINSTEPS, chicago, MESA Press.

Puskulluoglu M, Tomaszewski KA, Bottomley A, et al (2014). Validation of the polish version of the EORTC QLQ-BM22 module for the assessment of health-related quality of life in patients with bone metastases. *Qual Life Res*, **23**, 527-32.

Scott NW, Fayers PM, Aaronson NK, et al (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *J Clin Epidemiol*, **62**, 288-95.

Scott NW, Fayers PM, Aaronson NK, et al (2010). Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health Qual Life Outcomes*, **8**, 81.

Suhonen R, Schmidt LA, Katajisto J, et al (2013). Cross-cultural validity of the individualised care scale - a rasch model analysis. *J Clin Nurs*, **22**, 648-60.

Yekaninejad MS, Ahmadzadeh A, Mosavi SH, et al (2014). The reliability and validity of the iranian version of the european organization for research and treatment of cancer quality of life questionnaire for patients with bone metastases: the EORTC QLQ-BM22. *Expert Rev Pharmacoecon Outcomes Res*, **14**, 147-56.

Zeng L, Chow E, Bedard G, et al (2012). Quality of life after palliative radiation therapy for patients with painful bone metastases: results of an international study validating the EORTC QLQ-BM22. *Int J Radiat Oncol Biol Phys*, **84**, 337-42.

100.0

6.3

75.0

56.3

50.0

25.0

31.3

0

Newly diagnosed without treatment