Journal of the Formosan Medical Association (2019) 118, 215–222

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfma-online.com

Original Article

Psychometric evaluation of the WHOQOL-BREF, Taiwan version, across five kinds of Taiwanese cancer survivors: Rasch analysis and confirmatory factor analysis

Chung-Ying Lin [a], Jing-Shiang Hwang [b], Wen-Chung Wang [c,†], Wu-Wei Lai [d], Wu-Chou Su [e], Tzu-Yi Wu [f,g], Grace Yao [h,**], Jung-Der Wang [e,g,i,*]

[a] *Department of Rehabilitation Sciences, Faculty of Social Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*
[b] *Institute of Statistical Science, Academia Sinica, Taipei, Taiwan*
[c] *Department of Psychological Studies, The Education University of Hong Kong, Hong Kong*
[d] *Department of Surgery, National Cheng Kung University College of Medicine and Hospital, Taiwan*
[e] *Department of Internal Medicine, National Cheng Kung University College of Medicine and Hospital, Tainan, Taiwan*
[f] *Institute of Economics, Academia Sinica, Taipei, Taiwan*
[g] *Department of Public Health, College of Medicine, National Cheng Kung University Hospital, National Cheng Kung University, Tainan, Taiwan*
[h] *Department of Psychology, College of Science, National Taiwan University, Taipei, Taiwan*
[i] *Department of Occupational and Environmental Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan*

Received 20 March 2018; accepted 28 March 2018

KEYWORDS
Confirmatory factor analysis;
Cancer;
Psychometric properties;
Quality of life;
Rasch analysis

Background: Quality of life (QoL) is important for clinicians to evaluate how cancer survivors judge their sense of well-being, and WHOQOL-BREF may be a good tool for clinical use. However, at least three issues remain unresolved: (1) the psychometric properties of the WHOQOL-BREF for cancer patients are insufficient; (2) the scoring method used for WHOQOL-BREF needs to be clarify; (3) whether different types of cancer patients interpret the WHOQOL-BREF similarly.
Methods: We recruited 1000 outpatients with head/neck cancer, 1000 with colorectal cancer, 965 with liver cancer, 1438 with lung cancer and 1299 with gynecologic cancers in a medical center. Data analyses included Rasch models, confirmatory factor analysis (CFA), and Pearson correlations.

* Corresponding author. Department of Public Health, College of Medicine, National Cheng Kung University, 1 University Road, Tainan 70101, Taiwan; Fax: +886 6 235 9033.
** Corresponding author. Department of Psychology, College of Science, National Taiwan University, Taipei, Taiwan.
E-mail addresses: kaiping@ntu.edu.tw (G. Yao), jdwang121@gmail.com (J.-D. Wang).
† Deceased February 15th, 2018.

https://doi.org/10.1016/j.jfma.2018.03.018

Results: The mean WHOQOL-BREF domain scores were between 13.34 and 14.77 among all participants. CFA supported construct validity; Rasch models revealed that almost all items were embedded in their expected domains and were interpreted similarly across five types of cancer patients; all correlation coefficients between Rasch scores and original domain scores were above 0.9.

Conclusion: The linear relationship between Rasch scores and domain scores suggested that the current calculations for domain scores were applicable and without serious bias. Clinical practitioners may regularly collect and record the WHOQOL-BREF domain scores into electronic health records.

Introduction

Cancer is one of the major causes of death and negatively affects patients' quality of life (QoL). However, most cancers of different organ-systems have become chronic, as they would usually be able to survive for more than 6 months.[1] If an individual suffering from cancer is still alive at the time of observation, he/she is defined as a cancer survivor. There were 15.5 million cancer survivors in the United States in 2016, and this number is expected to grow to 20.3 million by 2026.[2] Being diagnosed with cancer is a heavy burden for patients, because they will need to receive long-term follow-up and treatment, with the details varying according to the different stages and types of cancer. In addition to basic survival, patients' QoL is an important factor that can enable both clinicians and patients to understand whether a cancer survivor is living well, and the concept of QoL fits into the patient-centered care model that is now being advocated.[3]

QoL is viewed as a key outcome measure for cancer survivors.[4] Moreover, the European Organization for Research and Treatment of Cancer (EORTC) recognizes the significance of QoL and has developed cancer-specific QoL instruments for clinical use.[5] Although these cancer-specific QoL instruments help clinicians understand the patients' QoL specifically as it relates to their cancer symptoms, generic QoL instruments provide a direct comparison for people with different illnesses and health states[6] and would be useful for health policy decision. There is thus also a need to validate a generic QoL instrument for cancer care, and the World Health Organization Quality of Life Scale Brief Version (WHOQOL-BREF) is a good candidate for validation.

Although the psychometric properties of WHOQOL-BREF have been examined for various populations,[7,8] the results of these studies may not be applicable to cancer patients, because different populations have varied clinical characteristics. Moreover, a sound instrument needs to be tested using different statistical methods across different populations, because the nature of scientific inquiry is to accumulate evidence using different methods. The evidence of psychometric properties is highly dependent on the tested populations, and the reliability and validity may not be the same across different populations.[9] It is thus crucial for the WHOQOL-BREF to be tested for its psychometric properties in relation to cancer patients.

To the best of our knowledge, only three articles[10–12] examined the psychometric properties of WHOQOL-BREF; another three[4,13,14] examined the psychometric properties of the long version of WHOQOL-BREF, WHOQOL-100, for cancer patients. But some gaps in the literature remain to be addressed. First, previous studies testing the WHOQOL-BREF only focus on one type of cancer patient. Therefore, we have no information regarding the psychometric properties of the instrument across different types of cancer patients. Second, most studies used classical test theory (CTT) for testing, rather than using modern test theory, such as a Rasch model. This study intended to apply the Rasch model for psychometric assessment because it has the following advantages: (1) it converts the raw score into an interval scale (i.e., Rasch score with the unit of *logit*); (2) it separates the estimates of item difficulty and person ability; (3) it tests the item properties, and (4) it examines the threshold order of the response categories.[7] Moreover, given that the descriptors of WHOQOL-BREF are carefully selected, would the original domain scores be close to an interval scale like a Rasch score? Although Rasch scores seem to be more normally distributed and would be more sensitive to detect differences than the use of raw scores,[15] no study has investigated the linear relationship between the Rasch and raw scores. If the two types of scores are linearly correlated, or almost linearly correlated, the use of raw or original domain scores can somewhat be justified. Thus, the study purpose was to test the psychometric properties of the WHOQOL-BREF scores among five common types of cancer patients.

Methods

Participants

The Institutional Review Board of National Cheng Kung University Hospital approved this study, and all participants ($n = 5702$) voluntarily agreed to participate in the study with every one providing a written informed consent. The participants were recruited from the cancer centers of National Cheng Kung University Hospital between January 2012 and December 2014, and they all met the following inclusion criteria: (1) had a diagnosis of cancer (including head and neck cancer, colorectal cancer, liver cancer, lung cancer, and gynecologic cancers) corroborated by histopathology and/or

cytology; (2) were aware of their cancer diagnosis; (3) did not have any other malignancy; namely, patients with more than one cancer were excluded; and (4) had the ability to understand and complete the WHOQOL-BREF. In addition, we set the subject to item ratio at 35:1 to determine the sample size for each cancer type. Given that 26 items in the WHOQOL-BREF were used for analysis, each cancer type should have at least included 910 participants.

Instrument

The WHOQOL-BREF Taiwan version was developed following a standard translation procedure and psychometric evaluations.[7,16] The design of the Taiwanese version added 2 additional items to account for Taiwanese cultural adaptations, which resulted in a total of 28 items. In addition to 2 generic items, the other 26 items are distributed into Physical (7 items), Psychological (6 items), Social (4 items including a cultural adaptation item So4), and Environment (9 items including a cultural adaptation item En9) domains.

Item scores range from 1 (the worst condition) to 5 (the best condition), except for 3 items (Ph1, Ph2, and Ps6), which are reverse coded. Two scales (0—100 and 4—20; a higher score indicates better QoL) can be transformed into domain scores for each domain,[7,16] and we used the 4—20 scale in this study.

Data analysis

The descriptive analyses were conducted using SPSS 15.0 (SPSS Inc., Chicago, IL); the CTT analyses for psychometric properties were analyzed using R software (including psych and lavaan); the Rasch analyses were analyzed using WINSTEPS 3.75.0 (Winsteps.com, Chicago, IL).

CTT analyses

Internal consistency, concurrent validity, and confirmatory factor analysis (CFA) were the CTT analyses used to examine the psychometric properties of the WHOQOL-BREF. Internal consistency was tested using Cronbach's α, and a value of α larger than 0.7 indicates satisfactory internal consistency. Concurrent validity was tested using the correlations between two generic items on WHOQOL-BREF ("How would you rate your quality of life?" and "How satisfied are you with your health?") and each domain score (Physical, Psychological, Social, and Environment), and we anticipated moderate correlations (i.e., $r > 0.3$). CFA was used to test the four-correlated-factor structure of the WHOQOL-BREF, and we used diagonally weighted least square (DWLS) estimator to tackle the Likert-type scale in the WHOQOL-BREF. Three fit indices were used to determine whether the factor structure was supported: comparative fit index (CFI) > 0.9, root mean square of error approximation (RMSEA) < 0.08, and standardized root mean square residual (SRMR) < 0.08 suggest acceptable fit.[17]

Rasch analysis for psychometric evaluations

All of the Rasch models were analyzed using rating scale models (RSMs) because we assume that the difficulties of the descriptors were similar across items. The assumption is reasonable because the descriptors of the WHOQOL-BREF Taiwan version followed the guidelines of choices of descriptors very carefully, and evidence showed the similar difficulties across items.[16] The item properties in each domain were tested for different cancer types, and a total of 20 Rasch RSMs (4 WHOQOL-BREF domain $\times$ 5 cancer types) were conducted. Item fit statistics, including infit mean square (MnSq) and outfit MnSq, were used to examine whether an item fit well in its embedded domain. An acceptable range of MnSq is between 0.5—1.5,[18] and a higher MnSq value of an item indicates that the item is underfitted (i.e., unpredictable and/or outside of the concept); a lower MnSq value indicates that it is overfitted (i.e., there is some redundancy).[18,19]

The reproducibility of the hierarchical item difficulty was examined using item separation reliability, and reproducibility of the person ordering was examined using person separation reliability. In addition, the item separation index tested for the classification of the items, and the person separation index tested for the classifications of the persons. Acceptable values are 0.7 for reliability and 2 for the separation index.[20] Moreover, the difficulty of each item was presented using *logit* unit for each cancer type.

The ordering of the response categories was tested using average and step measures. The average measure examined the average QoL level of all patients who chose the particular category, and the step measure was the threshold between the response categories.[7] The monotonically increasing average and step measures indicate that the response categories for each item were successfully located in their expected order.

Differential item functioning (DIF) analysis was conducted to examine the invariance of the item-difficulty hierarchy across the five cancer types. If an item displays DIF, then its difficulty varies across different subgroups; that is, the subgroups interpret the item in different ways.[21] We applied a DIF contrast (i.e., *logit* of Group 1 minus *logit* of Group 2) > 0.5 *logit* as a substantial DIF.[21]

Association between original domain scores and Rasch scores

We first examine whether the Rasch score outperforms original domain score, and whether the Rasch score has the ability to be a criterion for the domain score. The coefficients of variation (CV; i.e., SD divided by the mean) were obtained to detect the differences in variability between Rasch and domain scores. The CV is a standardized measure that describes the dispersion of a probability distribution or a frequency distribution. The CV can thus help us compare the variations in different score ranges, and a greater value suggests a larger variation. Therefore, the larger one is more sensitive and has a greater capability to detect individual differences. In addition, analyses of variance (ANOVAs) with Bonferroni adjustment were applied to all Rasch scores and all domain scores to compare the differences in QoL across five types of cancer patients.

Results

A total of 5702 cancer patients participated in the study, and Table 1 summarizes their demographic details. CTT

Table 1 Demographics across five cancers patients.

	n (%)					χ^2 or (F)	Multiple comparisons[b]
	1. Head/Neck cancer	2. Colorectal cancer	3. Liver cancer	4. Lung cancer	5. Gynecologic cancers		
Total participants	1000 (100.0%)	1000 (100.0%)	965 (100.0%)	1438 (100.0%)	1299 (100.0%)	–	–
Age (yr)[a]	55.89 (11.38)	63.39 (12.46)	61.68 (11.22)	63.89 (11.80)	52.68 (12.87)	(209.62)	2,4 > 3 > 1 > 5
Gender (male)	859 (85.9%)	573 (57.3%)	711 (73.7%)	765 (53.2%)	0 (0.0%)	2056.69	–
Educational level						189.72	–
≤ elementary	290 (29.0%)	399 (39.9%)	419 (43.4%)	627 (43.6%)	398 (30.6%)		
Junior high	253 (25.3%)	158 (15.8%)	174 (18.0%)	209 (14.5%)	204 (15.7%)		
Senior high	278 (27.8%)	191 (19.1%)	204 (21.1%)	299 (20.8%)	398 (30.6%)		
≥ college	176 (17.6%)	230 (23.0%)	157 (16.3%)	207 (14.4%)	293 (22.5%)		
Marital status						72.83	–
Single	108 (10.8%)	67 (6.7%)	55 (5.7%)	73 (5.1%)	135 (10.4%)		
Currently married	747 (74.7%)	721 (72.1%)	723 (74.9%)	978 (68.0%)	861 (66.3%)		
Others	130 (13.0%)	187 (18.7%)	166 (17.2%)	274 (19.1%)	277 (21.3%)		
Current smoker (yes)	256 (25.6%)	88 (8.8%)	162 (16.8%)	130 (9.0%)	79 (6.1%)		
Current drinker (yes)	168 (16.8%)	100 (10.0%)	58 (6.0%)	144 (10.0%)	116 (8.9%)		

[a] Reported in Means (SD).
[b] Using Bonferroni correction.

results showed satisfactory internal consistency across all five types of cancer patients in the Physical ($\alpha = 0.794$ to 0.834), Psychological ($\alpha = 0.781$ to 0.827), and Environment domains ($\alpha = 0.730$ to 0.803), but unsatisfactory internal consistency in the Social domain ($\alpha = 0.487$ to 0.657). The concurrent validity of the WHOQOL-BREF was supported by the significantly moderate correlations between overall QoL and all domain scores ($r = 0.30$ to 0.60); between general health and all domain scores ($r = 0.31$ to 0.68). Moreover, the fit indices of the CFA models supported the factor structure in all five types of cancer patients: CFI = 0.991, RMSEA = 0.030, and SRMR = 0.045 for head and neck cancer; CFI = 0.974, RMSEA = 0.041, and SRMR = 0.056 for colorectal cancer; CFI = 0.981, RMSEA = 0.040, and SRMR = 0.055 for liver cancer; CFI = 0.981, RMSEA = 0.042, and SRMR = 0.055 for lung cancer; CFI = 0.984, RMSEA = 0.038, and SRMR = 0.051 for gynecologic cancers.

The misfit items identified by the Rasch models were almost the same across the five types of cancer survivors. Five items from the Physical (3 items), Psychological (1 item), and Environment (1 item) domains were misfit, while all the other items fit well in their embedded domains. Of the 5 misfit items, items Ph2 (*Medication*), Ps6 (*Negative feelings*), and En9 (*Eating*) were underfit for all types of cancer; item Ph1 (*Pain and discomfort*) was underfit for all cancers except for lung cancer; item Ph6 was overfit for colorectal cancer (Table 2).

Person separation and person reliability seemed to be adequate in all domains, except the Social domain. Nonetheless, all domains exhibited excellent item separation and item reliability (Appendix Table S1). The difficulties of the response categories (examined using an average measure) and those of the thresholds (examined using a step measure) increased monotonically in their expected orders (Appendix Table S2). In addition, the difficulties of the same items across the five types of cancer survivors were similar in all four WHOQOL-BREF domains, indicating a successful choice of descriptors. However, our visualized diagram (Fig. 1) shows that the following items had divergent difficulties across the five types of cancer survivors: Ph2 (*Medication*), Ps6 (*Negative feelings*), So4 (*Being respected*), and En9 (*Eating*).

Additional DIF tests corresponding to Fig. 1: item Ph2 (*Medication*) displayed substantial DIF in 6 comparisons between different types of cancers; Ps6 (*Negative feelings*) in 4 comparisons; So4 (*Being respected*) in 4 comparisons; En9 (*Eating*) in 4 comparisons (Appendix Table S3).

The domain scores calculated from the original item scores had smaller variability as compared with the converted Rasch scores in all domains across the five kinds of cancers (Appendix Table S4). Moreover, the WHOQOL-BREF score comparisons using the original domain scores across the five cancers were similar to those using the Rasch scores, though the Rasch scores seemed to have larger F values than did the original scores (Appendix Table S4). The linear correlations between the Rasch scores and the original domain scores were highly correlated: $r = 0.911–0.970$ ($p < 0.001$) for head/neck cancer; $= 0.910–0.962$ ($p < 0.001$) for colorectal cancer; $= 0.908–0.968$ ($p < 0.001$) for liver cancer; $= 0.922–0.967$ ($p < 0.001$) for lung cancer; $= 0.922–0.965$ ($p < 0.001$) for gynecologic cancers.

Discussion

The strengths of this study are as follows: (1) the participants included different major types of cancer survivors; (2) there was a large sample size for each type of cancer; (3) different statistical methods were used to examine the psychometric properties of the WHOQOL-BREF; (4) the ordering of the categories and DIF for cancer patients were tested; and (5) the evidence was provided with regard to using the original domain scores of the WHOQOL-BREF in clinical settings. In addition, results derived from CFA and

Table 2 Item difficulty, Rasch fit statistics for each item across different cancer survivors.

	Head and neck cancer			Colorectal cancer			Liver cancer			Lung cancer			Gynecologic cancers		
	D	Infit	Outfit	D	Infit	Outfit	D	Infit	Outfit	D	Infit	Outfit	D	Infit	Outfit
Phy															
Ph1	−1.26	**1.59**	1.48	−1.46	**1.62**	**1.53**	−1.45	**1.52**	**1.50**	−1.30	1.42	1.37	−1.21	**1.54**	1.49
Ph2	−0.75	**2.10**	**2.05**	−0.06	**1.98**	**1.96**	0.14	**2.02**	**2.03**	−0.07	**1.95**	**1.96**	−1.20	**1.95**	**1.84**
Ph3	0.91	0.70	0.70	0.70	0.73	0.73	0.71	0.64	0.64	0.64	0.71	0.72	1.05	0.76	0.76
Ph4	−0.12	0.85	0.79	−0.02	0.81	0.77	−0.09	0.78	0.74	0.12	0.79	0.77	−0.08	0.93	0.89
Ph5	0.82	1.05	1.05	0.77	1.07	1.07	0.78	1.04	1.06	0.61	1.14	1.17	1.31	1.09	1.10
Ph6	−0.03	0.54	0.51	−0.19	**0.48**	**0.47**	−0.26	0.54	0.51	−0.21	**0.49**	**0.47**	−0.06	0.52	0.50
Ph7	0.43	0.65	0.63	0.26	0.59	0.58	0.18	0.64	0.63	0.20	0.63	0.63	0.19	0.64	0.63
Psy															
Ps1	1.44	0.97	0.95	1.36	0.95	0.96	1.38	1.04	1.01	1.32	1.04	1.02	1.48	1.11	1.12
Ps2	−0.22	1.01	0.96	−0.21	0.99	0.94	−0.20	1.06	1.01	−0.21	1.06	1.00	−0.35	0.98	0.94
Ps3	0.29	0.81	0.79	0.37	0.87	0.83	0.44	0.83	0.82	0.36	0.85	0.84	0.50	0.86	0.85
Ps4	−0.13	1.11	1.07	−0.25	0.93	0.89	−0.34	0.98	0.94	−0.36	0.99	0.94	−0.60	1.00	0.97
Ps5	−0.25	0.62	0.61	−0.19	0.63	0.60	−0.21	0.63	0.60	−0.16	0.62	0.60	−0.48	0.60	0.59
Ps6	−1.13	**1.63**	**1.64**	−1.08	**1.85**	**1.87**	−1.08	**1.63**	**1.65**	−0.95	**1.58**	**1.60**	−0.56	**1.51**	**1.56**
Soc															
So1	−0.05	0.81	0.71	0.00	0.73	0.63	0.00	0.85	0.75	0.01	0.82	0.72	−0.15	0.87	0.78
So2	0.79	1.10	1.08	1.24	1.12	1.11	1.08	1.12	1.07	1.14	1.17	1.21	0.88	1.14	1.13
So3	−0.75	0.81	0.70	−0.66	0.80	0.68	−0.66	0.75	0.64	−0.80	0.82	0.70	−0.97	0.81	0.72
So4	0.00	1.30	1.18	−0.58	1.43	1.30	−0.42	1.30	1.21	−0.36	1.24	1.13	0.24	1.19	1.11
Env															
En1	0.00	0.83	0.79	0.12	0.83	0.80	0.07	0.87	0.81	0.27	0.94	0.88	0.41	0.88	0.86
En2	0.24	0.98	1.03	0.23	0.96	0.96	0.19	0.94	0.92	0.26	0.96	0.97	0.54	0.92	0.92
En3	1.20	0.93	1.00	1.10	0.86	0.89	1.25	0.93	0.97	1.25	0.95	0.98	1.20	1.00	1.05
En4	−0.09	0.87	0.83	0.13	0.92	0.85	0.06	0.90	0.84	0.00	0.88	0.80	−0.15	0.90	0.86
En5	0.69	1.16	1.19	0.86	1.26	1.27	0.85	1.33	1.38	0.99	1.31	1.32	1.08	1.30	1.33
En6	−0.53	0.61	0.59	−0.46	0.68	0.63	−0.46	0.58	0.54	−0.46	0.60	0.55	−0.27	0.74	0.71
En7	−0.88	0.72	0.71	−0.63	0.76	0.71	−0.59	0.85	0.83	−0.63	0.78	0.75	−0.67	0.77	0.76
En8	−0.69	0.62	0.57	−0.31	0.78	0.73	−0.34	0.70	0.65	−0.31	0.78	0.75	−0.53	0.68	0.65
En9	0.04	**1.98**	**2.13**	−1.04	**2.28**	**2.34**	−1.03	**2.13**	**2.23**	−1.37	**2.11**	**2.15**	−1.61	**1.96**	**1.99**

Reverse coded items are in *italics*; misfit values are in **bold**.
Phy = Physical; Psy = Psychological; Soc = Social; Env = Environment; D = Difficulty.

Rasch analysis all suggest that the WHOQOL-BREF had satisfactory construct validity but in different aspects. Specifically, CFA results indicated that the factorial structure of WHOQOL-BREF was four factors. Therefore, physical, psychological, social, and environment are different constructs. Rasch analysis additionally suggested that each factor was unidimensional: the items under the same domain of WHOQOL-BREF are grouped together to demonstrate the same construct.

Our CTT results agree with the findings of previous studies[8,11,12,16] that used CTT methods to test the psychometric properties of the WHOQOL-BREF. In addition, our finding of unsatisfactory internal consistency in the Social domain corresponded to the low Cronbach's α found in Van Esch et al.[12] and Oliveria et al.[11] We agree with the suggested explanation for this can probably be attributed to the small number of items in the Social domain. Another possible explanation for the low internal consistency may be the participants' characteristics as our Rasch analyses showed that person separation reliability was low (the values ranged from 0.57 to 0.69) in the Social domain, while the item separation reliability was high (values ranged from 0.99 to 1.00).

Studies using Rasch analysis found that the WHOQOL-BREF had no misfit items in the general population[22] or in community-dwelling elderly people.[23] In contrast, three misfit items (all the negatively worded items) were found in a heroin-dependent sample,[7] and one misfit item (Ph5: *sleep*) was found in people with depression.[24,25] One possible explanation for these diverse results for the misfit items is cultural differences, while an alternative explanation is the different health conditions in the different populations surveyed. If we only examine studies carried out within the same culture,[7,23,26] only Chang et al.[7] and our study found the negatively worded items to be misfit. Nevertheless, the fit statistics of the three negatively worded items were almost the highest in their embedded domains in Liang et al.[23] (Infit MnSq = 1.05 [Ph1: *pain and discomfort*] and 1.35 [Ph2: *medication*] vs. 0.72−0.93 and 1.20 [Ph5: *sleep*]; = 1.38 [Ps6: *negative feelings*] vs. 0.82−0.97) and Wang et al.[26] (Infit/Outfit MnSq = 1.19/1.27 [Ph1] and 1.18/1.30 [Ph2] vs. 0.83−1.02/0.81−1.02; = 1.15/1.17 [Ps6] vs. 0.87−1.10/0.86−1.10). Therefore, we tentatively conclude that the misfitting of the three negatively worded items can be at least partially explained by the trivial wording effects.

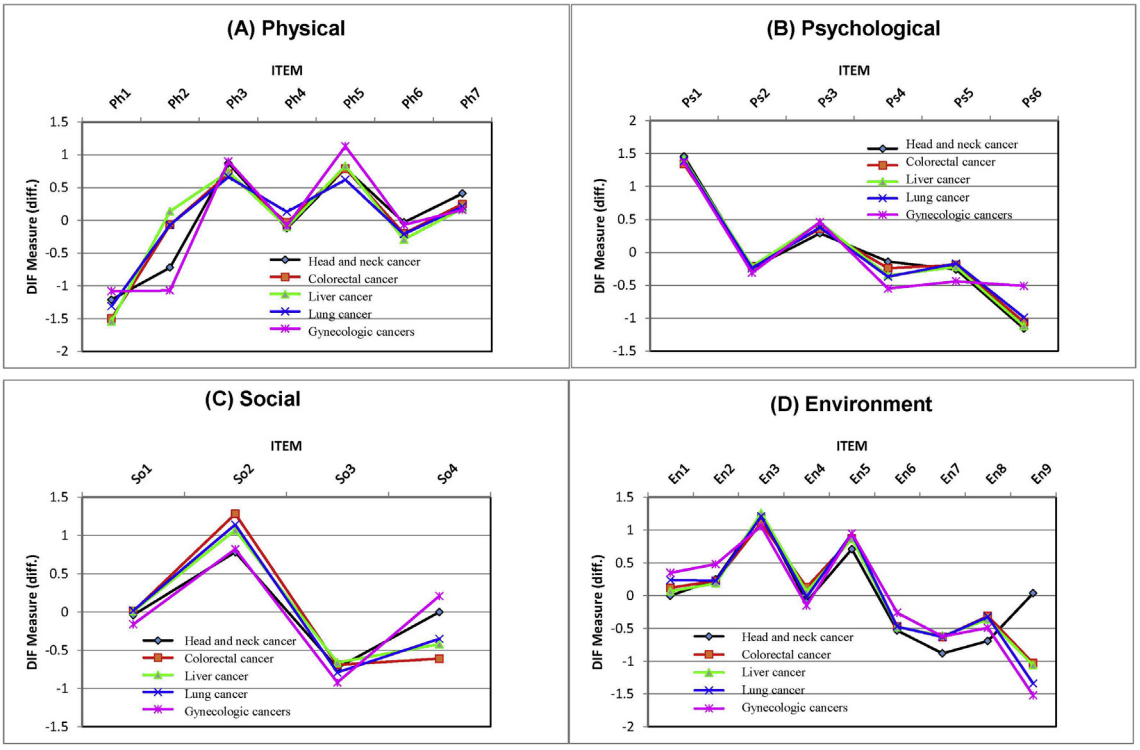

Figure 1 Differential item functioning (DIF) across five cancer groups.

Surprisingly, we found that the item En9 (*Eating*) was extremely misfitted (Infit/Outfit MnSq = 1.98–2.28/1.99–2.34). One potential explanation for this is that the Environment domain may not be the underlying construct of En9 although it is an important indicator measuring QoL for Taiwanese subjects.[16] Because both Chang et al.[7] and Liang et al.[23] found that En9 fit well in its embedded domain, we suspect that this item could be embedded in the Environment domain for populations other than cancer survivors. The major reason for this difference may be that the dietary guidelines and/or advice for cancer survivors are usually very strict, which would reduce their chances of dining outside the home and/or hospital. As such, cancer survivors may not think En9 is an item within the Environment domain.

In addition to the misfit items, some DIF items in the WHOQOL-BREF need to be taken into consideration when measuring patients' QoL. Specifically, (1) survivors of different types of cancer interpreted the *dependency on medication* (Ph2) differently; (2) patients with gynecologic cancers were more likely to express more *negative feelings* (Ps6) than those with head/neck cancer, colorectal cancer, and liver cancer; (3) survivors of gynecologic cancers and head/neck cancer felt that *being respected* (So4) was more difficult than other types of cancer patients did; (4) head/neck cancer patients considered *eating* (En9) to be more difficult than other types of cancer patients did.

A possible explanation for the different interpretations across different types of cancer survivors with regard to *dependency of medication* is that they received quite different treatments, which then had different side effects. Therefore, they may interpret their dependency on medications differently. It is reasonable for patients with gynecologic cancers to express more *negative feelings* as compared with other types of cancer survivors because women tend to express their emotions more than men do.[27] In addition, patients with gynecologic cancers and head/neck cancer suffer more with regard to their body image[17,28] as compared to other types of cancer survivors, and thus, may feel devalued and that it is difficult to be respected. Moreover, head and neck cancer patients with oropharyngeal cancer are more likely to have eating and drinking problems than are those with laryngeal cancer,[29] and thus are more likely to have problems with opening the mouth, eating, drinking, dry mouth, voice, and speech.

In addition, we found that the domain scores of the WHOQOL-BREF can be applied to clinical use because of the high correlations between the Rasch and domain scores, and because comparisons among different cancers show similar results for the Rasch and original domain scores. Our results agreed with previous findings[15] which suggested that using Rasch scores for ANOVA leads to the detection of larger F values than can obtained using the original domain scores. The larger F values also correspond to our CV results, in that the Rasch scores had larger CV values than did the domain scores. In other words, the Rasch scores had more capability to differentiate the performances of different individuals as compared with the original domain scores, and may have the sensitivity to detect minor changes. Also, the similarities between Rasch scores and domain scores in comparing different types of cancer survivors may be related to our large sample size. That is, Rasch scores and domains scores all had large F values. If our sample were in moderate size, Rasch scores may detect some significant findings that are not found using domain scores because of the large CV in Rasch scores. In order to corroborate this conjecture, we randomly selected half of

our participants to reanalyze the ANOVA across the same five cancers. Indeed, we found that Social domain became nonsignificant in domain score but retained significant in Rasch score (Appendix Table S5). Therefore, we also recommend the use of Rasch scores in the future if an information system which can automatically convert the raw scores into these is established. Nevertheless, the use of original domain scores seems justified based on our results, and may be a viable alternative for small hospitals/clinics to regularly collect and explore QoL data with patients in daily practice with minimal costs. In addition, our results derived from the domain scores demonstrated that WHOQOL-BREF had the ability to differentiate people with different types of cancer. Also, the sensitivity of item scores of the WHOQOL-BREF has been evidenced in a recent study on patients with lung cancer: WHOQOL-BREF detected the QOL differences in many item scores between two molecular target therapies.[30] As these adverse effects are not directly related to the respiratory system, they might not be picked up by usual lung-cancer specific items.

There are several limitations in this study. First, the participants were solely from the same medical center; our results may therefore have limited generalizability. Second, because the WHOQOL-BREF is a generic questionnaire, which does not contain any cancer-specific items, especially symptoms, for each type of cancer, it cannot reflect the detailed changes in QoL for cancer patients. Nonetheless, a generic questionnaire with good psychometric property may still be useful in detecting QoL changes for potential adverse effects of cancer treatment.[30] Third, all the participants had the ability to complete the WHOQOL-BREF; that is, none of them were too severe to respond to the questionnaire. In other words, our results cannot be applicable to cancer survivors with poor consciousness or inability to attend oncology clinic. Followed by the third limitation, the lack of information in the clinical characteristics of our participants may restrict the generalizability of our study results. We did not deliberately obtain the clinical characteristics (e.g., stage, severity, and treatment status) at the time of measurements. As all patients who visited oncology clinic at the NCKUH were invited and measured, the likelihood of a specific stage/severity and/or treatment would be low.

In conclusion, our results demonstrate that the WHOQOL-BREF is basically a sound instrument to measure QoL for cancer patients. However, substantial DIF was found for items Ph2 (*Medication*), Ps6 (*Negative feelings*), So4 (*Being respected*), and En9 (*Eating*), and thus clinicians should be cautious about using these for comparisons of different types of cancer survivors because of the possibility of different interpretations. We also suggest that QoL should be measured for cancer survivors repeatedly to capture the dynamic changes of their health conditions instead of simply applying a one-time measure. Future studies are needed to explore how to best use the WHOQOL-BREF for survivors with cancer of different organ-systems.

Sources of funding

The author(s) received partial support for this research by the Ministry of Science and Technology of Taiwan (MOST 106-2627-M-006-015). However, the data collection was funded by MOHW 104-TDU-B-211-124-003. The first author was supported by the project of Health Cloud Project, Academia Sinica.

Conflicts of interest

All of these funding and authors have no conflicts of interest.

Acknowledgement

We thanked all the participants, the program leaders (Drs. Helen Hai-Wen Chen, Jenq-Chang Lee, Yan-Shen Shan, Jiann-Hui Ou, Yih-Jyh Lin, Shih-Ming Huang, Jenn-Ren Hsiao, and Ya-Ming Cheng) who recruited the cancer patients in the cancer center, and the assistants who helped to collect the data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jfma.2018.03.018.

References

1. Wu T-Y, Chung C-H, Lin C-N, Hwang J-S, Wang J-D. Lifetime risks, loss of life expectancy, and healthcare expenditures for 19 cancers in Taiwan. *Clin Epidemiol* 2018 (in press).
2. National Cancer Institute (NCI), American Cancer Society. *Surveillance and health services research*. 2016. https://www.cancer.org/research/cancer-facts-statistics/survivor-facts-figures.html. [Accessed 21 March 2017].
3. Barry MJ, Edgman-Levitan S. Shared decision making — the pinnacle of patient-centered care. *N Engl J Med* 2012;**366**:780—1.
4. Paredes T, Simões MR, Canavarro MC. Psychometric properties of the World Health Organization Quality of Life Questionnaire (WHOQOL-100) in Portuguese patients with sarcoma. *Psychol Health Med* 2010;**15**:420—33.
5. Groenvold M, Klee MC, Sprangers MA, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient—observer agreement. *J Clin Epidemiol* 1997;**50**:441—50.
6. Lin C-Y, Luh W-M, Yang A-L, Su C-T, Wang J-D, Ma H-I. Psychometric properties and gender invariance of the Chinese version of the self-report pediatric quality of life inventory version 4.0: short form is acceptable. *Qual Life Res* 2012;**21**:177—82.
7. Chang K-C, Wang J-D, Tang H-P, Cheng C-M, Lin C-Y. Psychometric evaluation using Rasch analysis of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment: further item validation. *Health Qual Life Outcomes* 2014;**12**:148.
8. Su C-T, Ng H-S, Yang A-L, Lin C-Y. Psychometric evaluation of the short form 36 health survey (SF-36) and the World Health Organization Quality of Life scale brief version (WHOQOL-BREF) for patients with schizophrenia. *Psychol Assess* 2014;**26**:980—9.
9. Thompson B. Guidelines for authors. *Educ Psychol Meas* 1994;**54**:837—47.

10. Lin C-Y, Yang S-C, Lai W-W, Su W-C, Wang J-D. Rasch models suggested the satisfactory psychometric properties of the World Health Organization Quality of Life—brief among lung cancer patients. *J Health Psychol* 2017;**22**:397—408.
11. Oliveria IS, Costa LCM, Manzoni ACT, Cabral CMN. Assessment of the measurement properties of quality of life questionnaires in Brazilian women with breast cancer. *Braz J Phys Ther* 2014; **18**:372—83.
12. Van Esch L, Den Oudsten BL, De Vries J. The World Health Organization Quality of Life Instrument-short form (WHOQOL-BREF) in women with breast problems. *Int J Clin Health Psychol* 2011;**11**:5—22.
13. Den Oudsten BL, Van Heck GL, Van der Steeg AF, Roukema JA, De Vries J. The WHOQOL-100 has good psychometric properties in breast cancer patients. *J Clin Epidemiol* 2009;**62**:195—205.
14. Tazaki M, Nakane Y, Endo T, Kakikawa F, Kano K, Kawano H, et al. Results of a qualitative and field study using the WHOQOL instrument for cancer patients. *Jpn J Clin Oncol* 1998;**28**: 134—41.
15. Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care* 2004;**42**:I25—36.
16. Yao G, Chung C-W, Yu C-F, Wang J-D. Development and verification of validity and reliability of the WHOQOL-BREF Taiwan version. *J Formos Med Assoc* 2002;**101**:342—51.
17. Fang SY, Lin YC, Chen TC, Lin C-Y. Impact of marital coping on the relationship between body image and sexuality among breast cancer survivors. *Support Care Cancer* 2015;**23**: 2551—9.
18. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans* 1994;**8**:370.
19. Chang C-C, Su J-A, Tsai C-S, Yen C-F, Liu J-H, Lin C-Y. Rasch analysis suggested three unidimensional domains for Affiliate Stigma Scale: additional psychometric evaluation. *J Clin Epidemiol* 2015;**68**:674—83.
20. Wang W-C, Wilson M. Exploring local item dependence using a random-effects facet model. *Appl Psychol Meas* 2005;**4**: 296—318.
21. Shih C-L, Wang W-C. Differential item functioning detection using the multiple indicators, multiple causes method with a pure short anchor. *Appl Psychol Meas* 2009;**33**:184—99.
22. Krägeloh CU, Kersten P, Rex Billington D, Hsu P-H, Shepherd D, Landon J, et al. Validation of the WHOQOL-BREF quality of life questionnaire for general use in New Zealand: confirmatory factor analysis and Rasch analysis. *Qual Life Res* 2013;**22**: 1451—7.
23. Liang W-M, Chang C-H, Yeh Y-C, Shy H-Y, Chen H-W, Lin M-R. Psychometric evaluation of the WHOQOL-BREF in community-dwelling older people in Taiwan using Rasch analysis. *Qual Life Res* 2009;**18**:605—18.
24. Rocha NS, Fleck MP. Validity of the Brazilian version of WHOQOL-BREF in depressed patients using Rasch modelling. *Rev Saude Publica* 2009;**43**:147—53.
25. Rocha NS, Power MJ, Bushnell DM, Fleck MP. Cross-cultural evaluation of the WHOQOL-BREF domains in primary care depressed patients using Rasch analysis. *Med Decis Making* 2012;**32**:41—55.
26. Wang W-C, Yao G, Tasi Y-J, Wang J-D, Hsieh C-L. Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. *Qual Life Res* 2006;**15**:607—20.
27. Spalek K, Fastenrath M, Ackermann S, Auschra B, Coynel B, Frey J, et al. Sex-dependent dissociation between emotional appraisal and memory: a large-scale behavioral and fMRI study. *J Neurosci* 2015;**35**:920—35.
28. Fingeret MC, Vidrine DJ, Reece GP, Gillenwater AM, Gritz ER. Multidimensional analysis of body image concerns among newly diagnosed patients with oral cavity cancer. *Head Neck* 2010; **32**:301—9.
29. Weber C, Dommerich S, Pau HW, Kramp B. Limited mouth opening after primary therapy of head and neck cancer. *Oral Maxillofac Surg* 2010;**14**:169—73.
30. Yang SC, Lin CC, Lai WW, Chang SM, Hwang JS, Su WC, et al. Dynamic changes in quality of life after three first-line therapies for EGFR mutation-positive advanced non-small cell lung cancer. *Ther Adv Med Oncol* 2018;**11**:1—11.