J Rehabil Med 2013; 45: 873–880

ORIGINAL REPORT

RASCH ANALYSIS OF THE WHOQOL-BREF IN POST POLIO SYNDROME

Ian M. Pomeroy, Dphil[1], Alan Tennant, PhD[2] and Carolyn A. Young, MD, FRCP[1]

From the [1]Walton Centre NHS Foundation Trust and University of Liverpool, Liverpool and [2]Faculty of Medicine and Health, The University of Leeds, Leeds, UK

Objective: **The World Health Organisation quality of life abbreviated scale (WHOQOL-BREF) was developed as a measure of quality of life across 4 separate health domains; physical health, psychological, social relationships and environment. This study evaluated the validity of the WHOQOL-BREF in post-polio syndrome by testing it for fit against the Rasch model.**

Results: **The scale was posted to 319 volunteers, 271 (85%) completed the scale with a mean age of 66.7 years (standard deviation 8.15); 64% were female. The social relationships domain fitted the Rasch model (χ^2 $p=0.19$) but reliability was low ($\alpha=0.69$) and there were insufficient items to test the assumption of unidimensionality. Solutions were derived for physical health ($p=0.45$, t-test$=1.5\%$, $\alpha=0.67$), psychological ($p=0.19$, t-test$=4.9\%$, $\alpha=0.78$) and environment domains ($p=0.48$, t-test$=6.0\%$ - lower confidence interval 3.4%, $\alpha=0.80$) by accounting for local dependence and to cancel out differential item functioning. An overall measure of quality of life, which combined all 4 domains was validated ($p=0.80$, t-test$=4.6\%$, $\alpha=0.81$). A transformation table for this total score is provided.**

Conclusion: **The 4 domains of the WHOQOL-BREF provide valid measures of quality of life in post-polio syndrome. The summed score was more reliable and better targeted and can be used as an ordinal estimate of quality of life.**

Key words: Rasch analysis; WHOQOL-BRE; post-polio syndrome.

J Rehabil Med 2013; 45: 873–880

Correspondence address: Ian M. Pomeroy, Dphil, Walton Centre NHS Foundation Trust and University of Liverpool, U.K. E-mail: ian.pomeroy@thewaltoncentre.nhs.uk

Accepted Apr 3, 2013; Epub ahead of print Aug 8, 2013

INTRODUCTION

Acute poliomyelitis was endemic in the developed world until the 1960s. The United Kingdom entered an epidemic phase of polio between 1947 and 1966 (1), and there were more than 50,000 cases in the United States in 1952 alone with an estimated 250,000 U.S. polio survivors (2). More recently, wildtype poliovirus outbreaks were recorded in 22 countries of the developing world during 2009–2010 (3). Late effects of polio have been recognised as the post-polio syndrome (PPS) which has a prevalence of 28–31% in two studies conducted 37–67 years after acute polio infection (4, 5). PPS is characterised by the onset of new neuromuscular symptoms which developed at least 15 years after the initial infection. These symptoms include extensive fatigue, muscle weakness, muscle atrophy, pain and cold intolerance (6–9).

Muscle weakness, pain and fatigue may be disabling in certain areas of life and may affect independence, participation and quality of life (QoL) (10–12). In an analysis of the interplay between impairment and QoL for those with PPS, 53% of the variation in QoL was attributable to antecedent factors (13). The model supported the hypothesis that QoL was the outcome of a complex interplay between factors such as severity of impairment, antecedent variables, and health promoting behaviours. Such a model is often referred to as a bio-opsychosocial model (14), representing as it does the interplay between biological, psychological and social variables. The study used a QoL scale which had been developed for multiple sclerosis, without any validation for its use in PPS. However, exploration of the bio-psychosocial model in the context of PPS, such as that reported above, requires a reliable, valid, and ideally cross-culturally valid measure of QoL, preferably one which has a strong theoretical or conceptual basis. The World Health Organisation quality of life abbreviated scale (WHOQOL- BREF) was designed to meet these criteria.

During the validation of the WHOQOL-BREF, QoL was defined as "individuals' perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns." This definition views QoL as a subjective evaluation which is embedded in a cultural social and environmental context which focuses upon respondents perceived QoL. This can therefore provide a measure of QoL which is applicable across diagnoses and cross-culturally (15). Measurement of QoL in this way facilitates a holistic approach to health care which is important in a multidimensional disease such as PPS.

The WHOQOL was developed in 15 international, culturally and economically diverse centres (15) and has been found to have good psychometric properties of reliability, discriminate validity and construct validity when tested across 23 different countries (16). The original 100 item scale includes 4 items from each of 24 different facets of QoL as well as 4 additional items assessing general QoL and health status. Structural equational modelling subsequently showed that the items could be drawn together into 4 domains; physical health, psychological, social relationships and environment (17). The abbreviated version (WHOQOL-BREF) was developed as a

self-administered scale in 20 centres across 18 countries. This 26 item scale includes one question from each of the original 24 facets grouped into the 4 domains as well as two questions assessing general health and QoL (18).

The internal construct validity of the WHOQOL–BREF was examined through fit of its data to the Rasch measurement model (19). The process of Rasch analysis measurement is not dependent on the distribution of subjects' abilities, given that the data fit the model (20). The Rasch model makes the following assumptions:

- The likelihood of a subject affirming an item in a scale is a logistic function of the item's difficulty and the subject's ability in the trait being measured.
- The items of the scale measure a single construct and can be added together to create a valid score (unidimensionality).
- The relative difficulty of an item remains constant across all ability levels of respondents (invariance of items).
- The scoring structure for polytomous items is working as expected (ordered category thresholds).
- The scale does not behave differently across different subgroups in the population (differential item functioning).

Application of the Rasch model to the WHOQOL-BREF therefore provides a method of testing whether each of the 4 domains are unidimensional and whether it is appropriate to add item scores together to create a valid domain score. The model also allows us to test that the category ordering of each of the items works as expected and whether there is any item bias in the form of differential item functioning by age and sex in the PPS population. Rasch analysis also allows us to test whether manipulating the scale by adding together all 4 domains can provide a valid global measure of QoL. Where the assumptions of the Rasch model hold, the scale can be viewed as a unidimensional scale and the raw score can be transformed to an interval scale which can be analysed using more powerful parametric statistical analysis, given appropriate distributional assumptions.

This study aimed to apply the Rasch model to evaluate the 4 domains of the WHOQOL-BREF and to test whether the domains could be combined to create a single unidimensional measure of QoL in PPS.

METHODS

Participants

Patients were recruited through a supraregional clinic and from volunteers obtained through the British Polio Fellowship. All potential subjects were provided with an information pack and were asked to complete a consent form and a research pro forma which included a questionnaire concerning established diagnostic features of PPS taken from Halstead's revised definition in 1991 and affirmed by the European Federation of Neurological Societies (7, 9). Eligible subjects were required to confirm the following features:

1. A confirmed history of polio
2. Partial or fairly complete recovery after the acute episode
3. A period of at least 15 years with neurological and functional stability
4. New muscle weakness in muscles previously affected or unaffected
5. Extensive fatigue

Patients were excluded from the study if they had a concomitant serious medical or psychiatric condition or if they were not able to give informed consent. Patients who were physically unable to complete the questionnaire were permitted to use a scribe to report their answers on the form. Patients who met the inclusion and exclusion criteria were posted the English language version of the WHOQOL-BREF and were asked to complete and return the scale. The validity of the inclusion criteria were checked in this population by reviewing 33 sets of clinical notes of patients from the post polio clinic who were involved in patient interviews performed as part of a related study. Three of these patients did not meet the self-reported inclusion criteria for the definition of PPS used in this study. The discrepancy was attributed in one patient to confusion about the definition of neurological and functional stability and in another by confusion about the term "extensive" fatigue. The third patient had left the tick box for fatigue blank. All 3 patients had been diagnosed with PPS by a consultant neurologist and this was confirmed on retrospective review of the notes and at interview. Ethics committee approval for the study was obtained from the Research Ethics Committee of the investigating institution. Data input was checked in a random sample of 10% of questionnaires and showed a single error out of 260 questions (0.4%).

Rasch analysis

Rasch analysis was performed using RUMM 2030 software (21) through a systematic, iterative procedure outlined briefly here and described in more detail in a previous publication (22). A likelihood ratio test was performed to determine whether the rating scale version of the Rasch model or the partial credit version should be used. Cronbach's alpha and the person separation index were used to assess internal reliability, values of > 0.70 were taken as evidence of sufficient reliability. Chi squared statistics were used to assess the fit of items and persons to the Rasch model and to assess item-trait interaction as a measure of the overall fit of the scale. Evidence of deviation from the Rasch model was concluded if the p-value in this test was less than 5% after the application of a Bonferroni correction for multiple comparisons. The validity of the 5 category structure of each scale was assessed by examining the ordered set of response thresholds for each of the items. If disordered thresholds were present (i.e. the response items do not show a logical progression across the trait being measured) then category responses were collapsed to solve this problem. Unidimensionality was tested through a series of independent t-tests comparing person estimates from subtests of positive and negatively correlating items derived from the first residual component (23). This procedure requires that at least 12 category thresholds are present in each of the subtests being compared. Significant multidimensionality was noted to be present if the lower confidence interval (CI) of the number of significant tests at the 5% level was greater than 5%. Local response dependency (interdependence of items) was examined by constructing a residual item correlation matrix between all items once the influence of the Rasch factor had been removed (24). Where dependency is observed, items are combined into testlets (25) we provide a commentary on using testlets to accommodate local response dependency of items as an appendix to this paper. Evidence of differential item functioning (DIF) by age group or gender was performed by examining whether person factors (age and sex) affect the functioning of the scales. Likewise, where DIF was observed, items are combined into testlets to see if DIF cancels at the test level. Evidence of DIF was accepted if the p-value derived from ANOVA analysis was significant at the 5% level with a Bonferroni correction applied.

RESULTS

A summary of recruitment to the study is shown in Fig. 1. Three hundred and nineteen patients were considered eligible for the study and 271 (84.9%) completed and returned the questionnaire. The mean age of responders was 66.7 years (SD 8.15) and 64% were female. For the purposes of DIF analysis the sample was split into 4 approximately equal age groups; 31–61 years (68 subjects), 62–65 years (73 subjects), 66–72

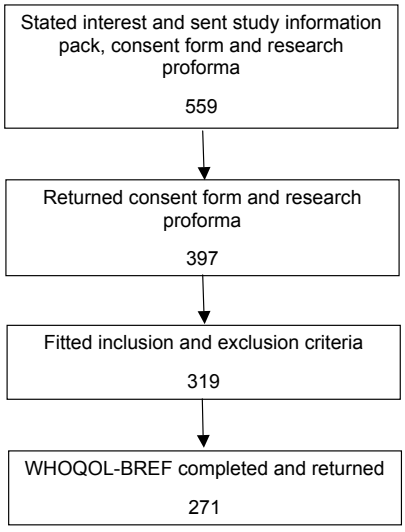


Fig. 1. Study recruitment.

years (61 subjects), 73–88 years (65 subjects), the age was not recorded in 4 subjects. Inspection of initial fit of the data to the Rasch model revealed that the likelihood ratio test was significant ($p < 0.0001$) and as a consequence the unrestricted parameterisation of the model (Partial Credit) was therefore adopted for all analyses. Fit statistics for original and revised subscales are shown in Table I, individual item fit statistics for original and revised subscales are shown in Tables II and III, respectively.

Physical health subscale

Initial analysis of the 6 item physical health subscale revealed good reliability, significant misfit to the Rasch model and significant multidimensionality (Table I, analysis 1). There was inter-item dependence seen between items 10, 17 and 18 and DIF by age group for item 15 ($p < 0.0001$) with younger persons with equivalent levels of QoL more likely to endorse the item at a higher level (how well are you able to get around). DIF by age group was seen in the opposite direction for item 16 (how satisfied are you with your sleep) with younger persons more likely to endorse the item although this did not

Table II. *Individual item fit statistics for the original subscales of the World Health Organisation quality of life abbreviated scale*

Scale/ Item	Location	Fit residual	χ^2	*p*
Physical				
3	–0.264	1.067	0.613	0.894
4	–0.637	2.483	3.001	0.392
10	0.061	–1.163	15.902	0.001
15	–0.415	1.585	4.048	0.256
16	–0.301	3.048	10.448	0.015
17	–0.329	–2.103	19.119	0.000
18	1.885	–0.669	6.083	0.101
Psychological				
5	–1.502	–1.008	6.678	0.083
6	–0.167	–1.806	7.771	0.051
7	–0.062	2.467	8.859	0.031
11	0.233	2.488	7.577	0.056
19	0.979	–2.004	6.232	0.101
26	0.519	0.336	0.587	0.899
Social				
20	–0.467	–0.521	6.793	0.079
21	1.031	0.929	2.940	0.401
22	–0.564	0.855	2.741	0.433
Environmental				
8	0.283	1.093	4.763	0.190
9	–0.370	–0.773	2.318	0.501
12	0.137	0.274	1.450	0.694
13	–0.446	–1.166	1.629	0.653
14	0.853	0.972	3.043	0.385
23	–0.469	0.655	2.128	0.546
24	0.215	0.398	1.461	0.691
25	–0.204	–0.242	4.439	0.218
Combined				
Physical	0.473	1.035	2.148	0.542
Psychological	–0.198	–0.881	6.509	0.089
Social	–0.116	1.633	0.381	0.944
Environmental	–0.159	–0.870	3.870	0.276

χ^2:Chi Squared; *p:* probability.

Table I. *Reliability and fit statistics for original and revised scales*

Analysis	Items	Reliability – alpha (PSI)	Overall fit	Item fit residual Mean (SD)	Person Fit residual Mean (SD)	% significant *t*-tests (lower CI)
Physical health						
1. Original scale	6	0.77 (0.75)	$\chi^2=59.2$ $p<0.0001$	0.61 (1.95)	–0.35 (1.34)	8.0 (5.3)
2. Combined items 15, 16, combined items 10, 17, 18	3	0.67 (0.70)	$\chi^2=12.0$ $p=0.447$	0.31 (2.31)	–0.45 (1.18)	1.5
Psychological health						
3. Original scale	7	0.81 (0.79)	$\chi^2=37.7$ $p=0.004$	0.08 (2.03)	–0.39(1.20)	7.4 (4.8)
4. Combined items 5, 6	6	0.78 (0.75)	$\chi^2=19.7$ $p=0.185$	–0.01 (1.23)	–0.40 (1.08)	4.9
Social relationships						
5. Original scale	3	0.62 (0.63)	$\chi^2=12.5$ $p=0.188$	0.42 (0.82)	–0.56 (1.23)	N/A
Environmental						
6. Original scale	8	0.81 (0.80)	$\chi^2=21.2$ ($p=0.63$)	0.19 (0.79)	–0.29 (1.07)	7.4 (5.3)
7. Combined items 12, 25	7	0.80 (0.80)	$\chi^2=20.7$ ($p=0.48$)	0.15 (1.10)	–0.33 (1.09)	6.0 (3.4)
Combined 4-domain scale						
8. Original scale	24	0.90 (0.90)	$\chi^2=182.9$ ($p<0.0001$)	0.43 (1.84)	–0.15 (1.45)	16.2 (13.6)
9. Items combined by subscale	4	0.80 (0.80)	$\chi^2=12.2$ ($p=0.531$)	0.15 (1.77)	–0.51 (1.05)	5.4 (2.8)
10. Combined physical and social subscales	3	0.85 (0.84)	$\chi^2=4.7$ ($p=0.856$)	0.13 (0.79)	–0.45 (0.93)	5.0 (2.3)

Table III. *Individual item fit statistics for the revised subscales of the World Health Organisation quality of life abbreviated scale*

Scale/Item	Location	Fit residual	χ^2	*p*
Physical				
3	0.090	1.291	1.960	0.581
4	–0.270	2.490	0.655	0.884
15,16	–0.143	0.071	6.294	0.098
10, 17, 18	0.324	–2.645	2.236	0.525
Psychological				
5, 6	–0.538	–2.051	2.531	0.470
7	–0.341	2.130	5.003	0.172
11	–0.037	1.958	2.770	0.428
19	0.668	–2.134	8.691	0.034
26	0.248	0.055	0.663	0.882
Environmental				
8	0.276	1.336	4.904	0.179
9	–0.372	–0.487	2.196	0.533
12, 25	–0.043	–1.430	5.492	0.139
13	–0.437	–0.960	1.375	0.712
14	0.853	1.179	3.253	0.354
23	–0.487	0.823	2.282	0.516
24	0.209	0.576	1.164	0.761
Combined				
Physical/Social	0.379	–0.787	0.259	0.967
Psychological	–0.198	0.476	2.521	0.474
Environmental	–0.181	0.677	1.708	0.635

χ^2: Chi Squared; *p*: probability.

reach significance when applying a Bonferroni correction ($p = 0.034$). There were no disordered thresholds. Combining items 15 and 16 into testlets cancelled out DIF by age group at the test level and combining items 10, 17 and 18 removed all positive residual correlations between items. The modified 3 testlet scale, which included all 6 original items (Table I, analysis 2) showed fit to the model and met the assumption of unidimensionality but reliability was low. The scale was well targeted in this population as only 1/271 patients were found to have a logit value outside of the range of the items (Fig. 2A).

Psychological subscale

Analysis of the psychological subscale showed good reliability but misfit to the model (Table I, analysis 3). There was a significant correlation between items 5 and 6 and DIF by gender for item 6 with women more likely to endorse the item (to what extent do you feel your life to be meaningful). Combining items 5 and 6 eliminated DIF and all positive correlations between the items. The modified 6 item scale (Table I, analysis 4) showed fit to the model with sufficient reliability. There was no significant misfit between individual persons or items, no disordered thresholds and no evidence of multidimensionality. The scale was well targeted in this population as only 3/271 patients were found to have a logit value outside of the range of the items (Fig. 2B).

Social relationships subscale

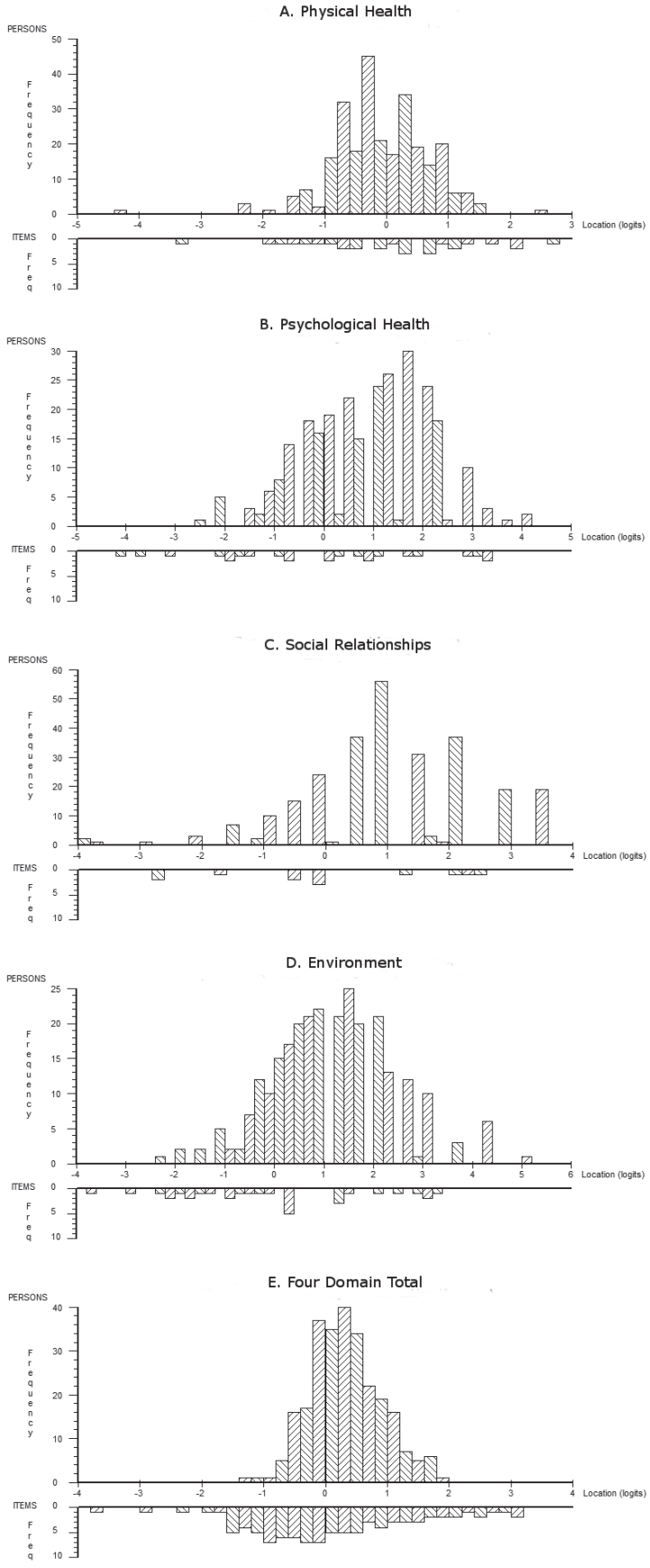
The 3-item social relationships subscale (Table I, analysis 5) showed fit to the model with no modifications required but had insufficient reliability. The subscale met the other assumptions of the Rasch model with no positive correlations between individual items, no DIF by age group or gender and no evidence of multidimensionality. The subscale had only 3 items and consequently only 9 response thresholds and it was therefore not possible to reliably assess for multidimensionality. In practice the scale had poor coverage across the social relationship construct, with large gaps between thresholds and a ceiling effect (Fig. 2C).

Environment subscale

Rasch analysis of the 8-item environmental subscale (Table I, analysis 6) revealed initial overall fit to the model but there was evidence of multidimensionality, disordered thresholds in item 25 and DIF by age group in item 12 ($p = 0.0002$) with older people more likely to endorse the item (have you enough money to meet your needs). Older people were less likely to endorse item 25 ($p = 0.0251$) (how satisfied are you with transport) although this did not reach significance with a Bonferroni correction applied. The scale was modified by combining responses 2 and 3 in item 25 and by combining items 12 and 25, which then showed no DIF for the new combined items when a Bonferroni correction for a 7-item scale was applied ($p = 0.0076$). Reanalysis of the modified scale (Table I, analysis 7) revealed fit to the model and reliability remained acceptable. There were no significant positive correlations greater than 0.1 between items and no evidence for multidimensionality. The scale was slightly off target in this population as 10/271 patients were found to have a logit value above that of the highest item (Fig. 2D).

Combined 4-domain scale

An attempt to validate a higher order measure of QoL related to physical, psychological, social and environmental factors was made by combining all items from the 4 subscales. An analysis of the unmodified 24 items in this scale (Table I, analysis 8) showed significant misfit to the model. There was significant multidimensionality present and assessment of inter-item dependence showed 10 pairs of items with correlations greater than 0.2, 10/11 of these correlations were between items from the same subscales. In view of this, the scale was reanalysed by combining items into 4 testlets grouped according to their original subscale. This modified 4 item scale (Table I, analysis 9) showed overall fit to the Rasch model but there was significant uniform DIF by gender in the physical health testlet with men scoring higher than women for an equivalent level of QoL ($p = 0.0033$). This uniform DIF by gender was shown to cancel out at the test level by combining the physical health and social relationships domains. Reanalysis of this 3 item scale (Table I, analysis 10) showed no evidence of multidimensionality (there were no positive correlations between any of the items, there was no evidence of misfit to the model and the solution showed good reliability. The total score for the combined scale showed strong correlations with item 1 which assessed patient's perception of overall QoL (Spearman's Rho = 0.64, $p < 0.00001$). The scale was very well targeted in this population as none of the patients were found to have a logit value outside of the range of the items (Fig. 2E).



These findings enable us to produce a nomogram to convert raw scores to transformed logit values, thereby converting the WHOQOL-BREF from an ordinal to a linear scale (Table IV). Since no items were removed or rescored, the raw score does not require any further transformation and is merely the sum of the scores for all individual items of the scale. Logit values corresponding to the raw scores obtained by subjects were determined using RUMM2030. The range of logit values was transformed back into the original range of 24–120 using a simple linear transformation.

DISCUSSION

This study has shown that the 4 domain structure of the WHOQOL-BREF is valid and it can be used to assess QoL in those with PPS. Each of the 4 subscales was shown to fit the Rasch model after accounting for local dependency and DIF. A total WHOQOL-BREF score has also been shown to be valid. The strategy of using testlets means that the WHOQOL-BREF can be used in its original format and summed to provide the subscale and total scores without deleting or rescoring items. As a consequence of this fit to the Rasch model, the WHOQOL-BREF scores can be transformed to interval-level latent estimates for use in parametric statistical methods, given distributional assumptions are met.

The application of these results is potentially limited by the fact that the study was performed in a single, developed country. However, the WHOQOL-BREF was designed to be applicable cross-culturally and further work is required to establish if the findings of this study are valid in different cultures, particularly in the developing world. Although the overall fit statistics were acceptable for each of the 4 subscales, other problems were highlighted by this analysis. Reliability was low in the social relationships and physical health domains and this is likely to be due to the short length of these subscales. The social relationships subscale consisted of only 3 items and the physical health subscale was reduced to 3 testlet-based items after items were combined to account for multidimensionality and DIF. DIF was seen in 3 items in this analysis although this was shown to cancel out

Fig. 2. Person–Item distribution plots comparing distribution of persons (upper plots) and items (lower plots) plotted on the same logit scales. The comparison between these two distributions illustrates the ability of each measure to cover the range of the latent trait (QoL) seen in the study population. Perfect matching of persons and items in a scale occurs when both distributions have a mean of zero logits and the distribution of persons does not fall outside the distribution of items.

Table IV. *Nomogram to convert raw scores to transformed values*

Raw score	Transformed value	Raw score	Transformed value
24	24.0	73	71.2
25	32.9	74	71.6
26	39.2	75	72.0
27	43.5	76	72.3
28	46.7	77	72.7
29	49.1	78	73.1
30	50.9	79	73.6
31	52.3	80	74.0
32	53.5	81	74.4
33	54.5	82	74.9
34	55.3	83	75.3
35	56.0	84	75.8
36	56.7	85	76.2
37	57.3	86	76.7
38	57.9	87	77.2
39	58.4	88	77.7
40	58.9	89	78.3
41	59.3	90	78.8
42	59.8	91	79.4
43	60.2	92	80.0
44	60.6	93	80.6
45	61.0	94	81.2
46	61.4	95	81.8
47	61.8	96	82.5
48	62.2	97	83.1
49	62.6	98	83.8
50	62.9	99	84.5
51	63.3	100	85.3
52	63.7	101	86.0
53	64.0	102	86.8
54	64.4	103	87.6
55	64.7	104	88.5
56	65.1	105	89.4
57	65.4	106	90.3
58	65.8	107	91.2
59	66.2	108	92.2
60	66.5	109	93.3
61	66.9	110	94.4
62	67.2	111	95.6
63	67.6	112	96.9
64	67.9	113	98.2
65	68.3	114	99.7
66	68.6	115	101.3
67	69.0	116	103.2
68	69.4	117	105.4
69	69.7	118	108.4
70	70.1	119	112.9
71	70.4	120	120.0
72	70.8		

This Table can be used to convert raw scores from the ordinal scale into transformed scores on a linear scale. This transformation is only valid where subjects have completed all items of the scale.

at the scale level in all cases. This suggests that DIF may be a problem in different populations and in different versions of the scale. The targeting of the subscales to the latent trait of QoL in this study was generally good, but the spread of item difficulty was restricted in the short social relationships domain and the environment subscale did not cover some of the higher scoring subjects. This suggests that there may be problems with appropriate targeting of the subscales in other populations.

Recent attempts at applying the Rasch model to the WHOQOL-BREF in patients with depression and in normal subjects have differed in their methodology and in their interpretation of fit to the model. Liang et al. (26) applied the Taiwanese version of the scale to the normal, elderly population and found all 4 domains to be unidimensional and reliable but 3 items showed DIF thereby suggesting a potential bias when using the scale in different groups. Rocha & Fleck (27) used the Brazilian version in patients with depression and also demonstrated appropriate fit to the model in all 4 domains but 2 items showed dependency and 4 items needed to be rescored.

Two papers have applied Rasch analysis to investigate whether the WHOQOL-BREF can be used as a single unidimensional construct. Noerholm et al. (28) found evidence of significant multidimensionality when applying the Danish version of the scale to the general population. Wang et al. (29) demonstrated that the scale could be made to fit the model but this required the deletion of 8 items due to substantial DIF, thereby detracting from the original construct validity of the scale. No previous research has found the WHOQOL-BREF to fit the Rasch model without significant rescoring or deletion of items. It is possible that the strategy of using testlets to overcome the problems of local dependency may be the significant difference between the findings of the current study, and previous studies, as has been shown elsewhere with other scales (30). When the items are grouped into 3 or 4 testlets, as in our Total Score solution, this also enacts the bi-factor solution, and it is important to note that the latent (person) estimate is based upon this dominant common factor.

Thus in this study, the scale was shown to meet the model's assumptions merely by combining items which were locally dependent. It did not require collapsing of thresholds, or deletion of items, and can therefore be administered in the field without modifications. This was reinforced by the finding of a strong correlation between the transformed scores and item one of the WHOQOL-BREF, which asked subjects to rate their overall QoL. Consequently the study provides a robust solution which overcame the problems of local dependency, and the targeting and reliability seen in analyses of the individual subscales, since the combined measures showed excellent reliability and a much wider coverage of the latent trait.

The summed 4-domain score is therefore a sufficient statistic and provides a simple ordinal estimate of QoL, reliable enough for individual use. In addition, where complete data is available for a subject, the raw scores can be converted using the nomogram into a linear measure of overall QoL in PPS thus providing a more appropriate quantitative outcome measure for service evaluation.

These findings show that the WHOQOL-BREF can be used to fulfil recommendations for further research in PPS which include measurement of QoL and the development of PPS-specific responsive, reliable and valid measures (2). The cross-cultural validity of the WHOQOL-BREF allow it to be used in the future for multicentre assessment of the efficacy of interventions in PPS, and to study further the factors that mediate the impact of health status upon QoL.

ACKNOWLEDGEMENTS

This work was supported by the British Polio Fellowship and the Neurological Disability Fund of the Walton Centre for Neurology and Neurosurgery. We would like to give special thanks to our participants who graciously gave their time to take part in this study.

REFERENCES

1. Hardy A. Poliomyelitis and the neurologists: the view from England, 1896–1966. Bull His Med 1997; 71: 249–272.
2. March of Dimes Steering Committee on Post-Polio Syndrome. March of Dimes International Conference on Post-Polio Syndrome: Identifying Best Practices in Diagnosis and Care. White Plains, N.Y, 2002. Available at: http://www.polioplace.org/sites/default/files/files/MOD-%20Identifying.pdf.
3. World Health Organisation. Weekly epidemiological record. Relevé épidémiologique hebdomadaire. N.16, 15 April 2011. 2011; 86: 153–160. Available at: http://www.who.int/wer/2011/wer8616.pdf.
4. Ramlow J, Alexander M, LaPorte R, Kaufmann C, Kuller L. Epidemiology of the post-polio syndrome. Am J Epidemiol 1992; 136: 769–786.
5. Ragonese P, Fierro B, Salemi G, Randisi G, Buffa D, D'Amelio M, et al. Prevalence and risk factors of post-polio syndrome in a cohort of polio survivors. J Neurol Sci 2005; 236: 31–35.
6. Dalakas MC. The post-polio syndrome as an evolved clinical entity. Definition and clinical description. Ann NY Acad Sci 1995; 753: 68–80.
7. Farbu E, Gilhus NE, Barnes MP, Borg K, de Visser M, Driessen A, et al. EFNS guideline on diagnosis and management of post-polio syndrome. Report of an EFNS task force. Eur J Neurol 2006; 13: 795–801.
8. Halstead LS, Rossi CD. Post-polio syndrome: clinical experience with 132 consecutive outpatients. Birth Defects Orig Artic Series 1987; 23: 13–26.
9. Halstead LS. Assessment and differential diagnosis for post-polio syndrome. Orthopedics 1991; 14: 1209–1217.
10. Kemp BJ, Krause JS. Depression and life satisfaction among people ageing with post-polio and spinal cord injury. Disabil Rehabil 1999; 21: 241–249.
11. Kling C, Persson A, Gardulf A. The health-related quality of life of patients suffering from the late effects of polio (post-polio). J Adv Nurs 2000; 32: 164–173.
12. Abresch RT, Carter GT, Jensen MP, Kilmer DD. Assessment of pain and health-related quality of life in slowly progressive neuromuscular disease. Am J Hosp Palliat Care 2002; 19: 39–48.
13. Stuifbergen AK, Seraphine A, Harrison T, Adachi E. An explanatory model of health promotion and quality of life for persons with post-polio syndrome. Soc Sci Med 2005; 60: 383–393.
14. Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1977; 196: 129–136.
15. Szabo S, Obot WG. The World Health quality of life WHOQOL assessment. In: Spilker B, editor. Quality of Life and Pharmacoeconomics in Clinical Trials. 2nd edn. Philadelphia: Lippincott-Raven; 1996, p. 335–362.
16. Skevington SM, Lotfy M, O'Connell KA; WHOQOL Group. 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.
17. WHOQOL-Group. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. Soc Sci Med 1998; 46: 1569–1585.
18. WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998; 28: 551–558.
19. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danish Institute for Educational Research; 1960.
20. Andrich D. Rasch models for measurement. London: Sage Publications; 1988.
21. RummLabs. Rumm2030. Australia: Rumm Laboratory PT Ltd; 2010.
22. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007; 57: 1358–1362.
23. Smith EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3: 205–231.
24. Andrich D MI. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. J Appl Meas 2008; 9: 200–215.
25. Wainer HK, G. Item clusters and computer adaptive testing: A case for testlets. J Educ Meas 1987; 24: 185–202.
26. Liang WM, Chang CH, Yeh YC, Shy HY, Chen HW, Lin MR. Psychometric evaluation of the WHOQOL-BREF in community-dwelling older people in Taiwan using Rasch analysis. Qual Life Res 2009; 18: 605–618.
27. Rocha NS, Fleck MP. Validity of the Brazilian version of WHOQOL-BREF in depressed patients using Rasch modelling. Rev Saude Publica 2009; 43: 147–153.
28. Noerholm V, Groenvold M, Watt T, Bjorner JB, Rasmussen NA, Bech P. Quality of life in the Danish general population – normative data and validity of WHOQOL-BREF using Rasch and item response theory models. Qual Life Res 2004; 13: 531–540.
29. Wang WC, Yao G, Tsai YJ, Wang JD, Hsieh CL. Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. Qual Life Res 2006; 15: 607–620.
30. Lundgren Nilsson Å, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM™) revisited. J Rehabil Med 2011; 43: 884–891.

APPENDIX I. Using testlets to accommodate local response dependency of items

In some scales, and particularly in health, it is not uncommon to find items that are locally dependent (1). For example, 'dressing upper body' with 'dressing lower body' (2). When data from these scales are fitted to the Rasch model, then it is possible that items such as these will have a high residual correlation (3). That is, they are correlated conditional upon the trait being measured, after the effect of the trait itself has been removed. This breaches one of the main assumptions underlying the summation of a set of items. The effect of this is to inflate reliability and cause misfit to the Rasch model, as the probability of response to the item is often much higher than expected by the model (4).

The issue is also not uncommon in certain educational tests where, for example, a comprehension test may have a single paragraph and several associated questions. This challenge to the local independence assumption has been dealt with by the use of testlets (5). A testlet is simply a summated set of items making a larger (super) item. It is a mechanism to deal with this problem, and does not affect the use, or scoring of the scale in any way. For example, the responses from 3 dichotomous walking items which ask about the distance walked can be added together to make one polytomous item, as if those questions had been asked as one question about how far a person can walk, with 3 response options reflecting distance (which may have been a better way to ask the question in the first place).

The presence of local (response) dependency is determined by examination of the residual correlation matrix. There is no definitive value of a correlation that indicates dependency, but it is usual to take 0.2 or 0.3 as indicative (6). With small scales it is worth finding out the mean residual correlation, as this may be negative, and thus a value of 0.1 may be indicative in these circumstances (this can be found out by exporting the matrix into Excel, removing the 'ones' on the diagonal, blocking, and Excel will report the mean). In some Rasch programmes such as RUMM, this process of creating testlets can be done as part of the general process of Rasch analysis (the subtest procedure in RUMM)(7). In other software, after the initial investigation, it may be necessary to group items together in the primary data file (e.g. SPSS) and to re-enter the data. Once this has been done, and the data are re-entered, then the testlets just become the same as any other polytomous item.

References

1. Wright B.D. Local dependency, correlations and principal components. Rasch Measurement Transactions 1996; 10: 509–511.
2. Lundgren Nilsson Å, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM™) revisited. J Rehabil Med. 2011; 43: 884–891.
3. Rasch, G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press, 1960.
4. Marais I, Andrich D. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. J Appl Meas 2008; 9: 200–215
5. Wainer H, Kiely G. Item clusters and computer adaptive testing: A case for testlets. J Educ Meas 1987; 24: 185–202.
6. Andrich D, Marais I. Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. J Appl Meas 2008; 9: 105–124.
7. Andrich, D, Sheridan BED, Luo, G. RUMM2030: Rasch unidimensional models for measurement. Perth, Western Australia: RUMM Laboratory, 2009.