Qual Life Res (2012) 21:237–246
DOI 10.1007/s11136-011-9936-1

The validity of subjective quality of life measures in psychotic patients with severe psychopathology and cognitive deficits: an item response model analysis

Ulrich Reininghaus · Rosemarie McCabe · Tom Burns · Tim Croudace · Stefan Priebe

Accepted: 17 May 2011 / Published online: 12 June 2011

Abstract
Purpose Subjective quality of life (SQOL) is an established patient-reported outcome in the evaluation of treatments for psychosis. The use of SQOL measures in the presence of psychiatric symptoms and cognitive deficits has been questioned. However, there is little evidence on whether items function differently as indicators of SQOL in psychotic patients with different levels of symptoms and deficits. Substantial differential item functioning (DIF) would, indeed, challenge the validity of established measures. We aimed to investigate the validity of a widely used measure of subjective quality of life (SQOL), i.e., the Lancashire Quality of Life Profile (LQOLP), in the presence of cognitive deficits and psychiatric symptoms in patients with severe and enduring psychosis.
Method We analysed SQOL ratings of 690 psychotic patients on the LQOLP using item response modelling to detect differential item functioning (DIF) attributable to psychiatric symptoms and cognitive deficits.
Results Patients with more severe general psychopathology were less likely to rate their 'personal safety' positively (OR .96, 95% CI .93–.99). More severely depressed patients were less likely to endorse positive 'life as a whole' (OR .93, 95% CI .89–.98) and 'mental health' (OR .93, 95% CI .91–.97) ratings. There was no DIF attributable to cognitive deficits.
Conclusions The findings suggest that the validity of the LQOLP in psychotic patients may be impaired by DIF due to psychopathology, although the magnitude of effects is unlikely to be of clinical significance. The validity appears not to be compromised by cognitive deficits.

Keywords Quality of life · Patient-reported outcomes · Psychosis · Validity · Differential item functioning · Item response theory

U. Reininghaus (✉) · R. McCabe · S. Priebe
Unit for Social and Community Psychiatry, Barts and the London School of Medicine, Newham Centre for Mental Health, Queen Mary University of London, London E13 8SP, UK
e-mail: u.reininghaus@qmul.ac.uk

U. Reininghaus
Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK

T. Burns
Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK

T. Croudace
Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK

Introduction

Subjective quality of life (SQOL) is an established patient-reported outcome (PRO) in the evaluation of treatment for patients with psychosis [1]. Self-report instruments are widely used to measure SQOL as a multi-domain concept using a conceptual framework, in which items (e.g. satisfaction with physical health) are grouped within domains (e.g. health), and domains within a more general SQOL concept [2, 3]. Both research studies and service providers now extensively draw on SQOL to assess treatment benefits for patients [4–7].

Cognitive deficits are increasingly considered as an important characteristic of psychosis [8]. There is, moreover, evidence that symptom profiles of patients with psychosis partition into positive and negative syndromes and, more recently, syndromes of disorganization, mania,

and depression [9–12]. Numerous studies have reported that both cognitive deficits [13–15] and psychiatric symptoms [16, 17] are associated with SQOL ratings. This has led several researchers to question the use of self-report measures of quality of life in the presence of psychiatric symptoms and cognitive deficits [18–21], with some proposing to entirely discard them [22–24]. Most concerns relate to affective and cognitive bias of subjective ratings. Katschnig [20] suggested that an 'affective fallacy' may contaminate ratings, so that depressed patients would consistently rate their life satisfaction as worse than it would appear for an independent observer. Atkinson et al. [18] proposed that emotional withdrawal and affective blunting may lead to higher SQOL ratings, as these symptoms may minimize the impact of illness-related dysfunction on subjective ratings of various life domains. Moreover, it has been suggested that subjective ratings might simply be wrong as patients may be unable to assess their life situation due to cognitive deficits [20]. In contrast, proponents of subjective measures emphasized that understanding the unique perspectives of psychotic patients is important, as it reflects their role as active partners rather than passive recipients of treatment [25–27]. In any case, potential problems in the measurement of SQOL in this patient group remain a challenge for psychosis research.

There is, however, only limited evidence on whether symptoms and cognitive deficits bias ratings at an item level [28, 29], that is, little is known as to whether items function differently as indicators of SQOL in patients with different levels of symptoms and deficits. For instance, patients with the same SQOL but varying levels of psychiatric symptoms or cognitive deficits may differ in endorsing lower or higher ratings on a SQOL item. Indeed, the presence of substantial differential item functioning (DIF) may pose a threat to the validity of SQOL measures. Specifically, it may impair their structural validity, i.e., the extent to which SQOL scores are an adequate reflection of the outcome to be measured [30], and therefore, represents an important source of evidence that may inform the use and development of SQOL measures in psychotic patients. Given the inverse relationship between psychiatric symptoms and SQOL that has been previously found by meta- and pooled analyses in patients with psychosis [16, 17], it is attractive to hypothesize that patients with the same SQOL but more severe symptoms are more likely to endorse more negative item responses on their satisfaction with life in general and specific life domains. Similarly, as previous studies have reported cognitive performance to be inversely related to SQOL in psychotic patients [13–15], it may be speculated that patients with lower cognitive performance (i.e. more deficits) are more likely to endorse higher SQOL ratings.

The Lancashire Quality of Life Profile (LQOLP) [31] is a widely used measure to assess SQOL in patients with severe and enduring psychosis in mental health service research in Europe [17, 32]. The psychometric properties of the LQOLP have been examined in a number of studies, based on principles of classical test theory [31–34]. Although widely used in psychotic patients, there is no evidence from item response model analysis on whether LQOLP ratings are affected by DIF due to psychiatric symptoms and cognitive deficits.

While evidence on DIF in PRO measures is emerging in various medical disorders [29, 35], there have been only a few studies considering this in psychiatric disorders [36, 37], fewer still in psychosis [5]. Psychiatric symptoms and cognitive deficits may bias the measurement of SQOL, but this has yet to be tested. We aimed to detect any DIF in SQOL ratings of patients with severe and enduring psychosis on a widely used measure of SQOL (i.e. the LQOLP) that can be attributed to psychiatric symptoms and cognitive deficits.

Method

Sample

We analysed data from the UK700 study [38], a multicentre randomized controlled trial comparing the effect of intensive case management (defined as smaller case-load size) in patients with severe and enduring psychosis. The data presented here are the SQOL assessments made at baseline. Patients were recruited from four inner-city mental health services in London and Manchester (UK). Approval for the study was obtained from the four local ethics committees and performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. The following inclusion criteria were used: (1) A primary diagnosis of psychosis according to the Research Diagnostic Criteria [39], (2) aged between 16 and 65 years, (3) hospitalized for psychotic symptoms at least twice within the past 2 years, (4) no obvious organic brain damage, (5) no primary diagnosis of substance abuse. During the study period, 708 patients were recruited from the four centres. Of these, 690 patients completed the questions on SQOL at baseline and were included into the current study. More detailed information on the UK700 study is available in Burns et al. [38].

Measures

Socio-demographic variables including age, gender, ethnicity, and duration of illness were collected from patients and case notes.

SQOL was measured using the LQOLP [31]. The LQOLP was based on Lehman's approach, operationalizing SQOL as satisfaction with life in general and in major life domains [2].

The LQOLP contains 24 items on patients' satisfaction with life in general and the following nine life domains: work, finances, social relations, leisure, living situation, safety, family relations, religion, and health. Each item is rated on a 7-point Life Satisfaction Scale (LSS).

General psychopathology was assessed using the Comprehensive Psychiatric Rating Scale (CPRS) [40], which rates the severity of 65 items of psychopathology over the preceding week on a 0–3 scale. Subscale scores for depression [Montgomery-Asberg Depression Rating Scale (MADRS)] [41] and positive symptoms were derived from the CPRS. Negative symptoms were measured on the Scale for the Assessment of Negative Symptoms (SANS) [42], which contains 24 items assessed over the same time frame as the CPRS. Higher (continuous) total CPRS, MADRS, or SANS scores indicate more severe symptomatology [40–42].

The National Adult Reading Test (NART) was used to measure premorbid intelligence [43]. The NART is an oral reading test consisting of 50 short and irregular words. Test performance has been reported to be relatively resistant to psychiatric disorder [44]. The NART error score converts into an estimate of pre-morbid fullscale, verbal, and performance IQ.

The Trail Making Test (TMT) was used to assess processing speed during attention and memory [45, 46]. The TMT consists of two parts, TMT-A and TMT-B. Task requirement for the TMT-A is to draw lines sequentially connecting 25 encircled numbers on a sheet of paper. TMT-B requires an individual to alternate between numbers and letters (e.g. 1, A, 2, B, 3, C, etc.). The score on each part is the time spent to complete the task. While TMT-A assesses right hemisphere integrity including visual scanning and spatial skills, TMT-B reflects left hemisphere function including direction of behaviour according to a complex plan as well as language and symbol manipulation.

Statistical analysis

To detect DIF attributable to psychiatric symptoms and cognitive deficits, analyses based on item response modelling were performed in MPlus, Version 5.2 [47], using multiple indicator multiple cause (MIMIC) models [48–50]. In line with Woods [50], the distinguishing feature of MIMIC models is that at least one observed variable, or covariate, predicts a latent variable. In order to detect DIF, MIMIC models additionally include an effect of a covariate on an item response [50]. There is good evidence on the equivalence of MIMIC models and DIF analysis in the IRT framework [51, 52]. MIMIC models offer a number of advantages over other approaches for detecting DIF [50]. These include the use of continuous covariates as well as the test of a direct effect of the covariate on the latent factor in the context of multidimensional models including the bifactor model [37, 51, 52]. As both cognitive deficits and psychiatric symptoms are most commonly operationalized as continuous variables and the LQOLP purports to measure SQOL as a multidomain concepts, MIMIC models were preferred over other techniques.

In line with Yang et al. [37], model estimation used both the robust weighted least squares means and variance adjusted estimator (WLSMV) and the full information maximum likelihood estimator with robust standard errors (MLR). Both WLSMV and MLR estimators return coefficients equivalent to item response theory (IRT) parameters extended to polytomous items. The two methods were employed to utilize their respective advantages for detecting DIF, that is, the WLSMV estimator and its fit statistics offer the advantage that convenient and efficient model building procedures have been delineated for the purpose of detecting DIF [37, 47–50]. By comparison, Yang et al. [37] noted that the MLR estimator offers the advantage of expressing regression coefficients as odds ratios as a basis for assessing the magnitude of DIF [50]. Data were assumed to be missing at random, which allowed for inclusion of the full sample using WLSMV and MLR estimators.

For models estimated using WLSMV, model fit was assessed using the root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker Lewis index (TLI) [53–55]. A good model fit is generally indicated by a low RMSEA (below .10 for acceptable and below .05 for very good fit) and a high comparative fit index (CFI) and Tucker Lewis index (TLI) (above .90 for acceptable and above .95 for very good fit) [51]. Overall fit of models estimated using MLR was examined using the log-likelihood, the Bayesian information criterion (BIC), and the Akaike information criterion (AIC) [56]. For these indices, lower values than for the comparison model indicate better model fit [57].

To establish the best fitting measurement model as a basis for performing DIF analysis, a unidimensional model with one general SQOL factor was compared with a bifactor model with one general SQOL factor and eight uncorrelated domain factors. The bifactor model recognizes that item responses depend both on a single general SQOL factor that explains covariance among all item responses and, independently, on domain factors that only account for responses to items of particular life domains [58–62]. Good evidence has been found for the bifactor model for Lehman's [63] Quality of Life Interview from which the LQOLP has been developed [59]. This model was compared with a standard unidimensional model to test the uni-versus multidimensional structure of the LQOLP item response data. Standardized factor loadings were computed to investigate the discriminative ability of items. In line with Reise et al. [62], factor loadings in the unidimensional model were compared

with those in the bifactor model to examine the amount of distortion that may occur when unidimensional models are fit to multidimensional data.

MIMIC models, in their extension to bifactor item response models by Yang et al. [37], were estimated to detect DIF using a stepwise model building procedure based on WLSMV [37, 52, 57]. First of all, general and specific effects of symptoms and cognitive deficits on latent general and domain factors were included into the model. Analyses were then adjusted for potential confounders previously found to be important for SQOL, i.e., age, gender, ethnicity, length of illness, and diagnosis [17, 64]. To detect DIF, modification indices (MI) were inspected to identify DIF effects of symptoms and deficits on item responses that, if freely estimated, would significantly improve model fit as indicated by a MI above 3.84. This step in the analysis involved a forward stepwise procedure to identify regression effects with the largest MI and, subsequently, freely estimating the parameter for this effect that had been constrained to be zero in the baseline model. A robust χ^2 model difference test was conducted to examine the statistical significance of DIF effects. The forward stepwise procedure continued until no significant effect was found [37, 57]. In a last step, the final model was re-estimated using MLR to obtain proportional odds ratios (ORs) as a basis for assessing the magnitude of DIF effects [37, 57, 65, 66]. In this model, a proportional OR of 2.0 is interpreted such that for a one unit increase in the independent variable (e.g. psychiatric symptoms or cognitive deficits), the odds of endorsing, e.g., the lowest (a rating of 1 = 'couldn't be worse' on the LQOLP) versus all higher categories (e.g. ratings of 2 'displeased' to 7 = 'couldn't be better' on the LQOLP) of the item affected by DIF is 2.0 times greater, as is the odds of endorsing the next lowest category (e.g. a rating of 2 = 'displeased' on the LQOLP) versus all higher categories (e.g. ratings of 3 = 'mostly dissatisfied' to 7 'could not be better' on the LQOLP). In line with Cole et al. [65], the magnitude was considered as large if proportional ORs were $\leq .5$ or ≥ 2.0. The impact of DIF effects was examined comparing the association of psychiatric symptoms and cognitive deficits with underlying SQOL (i.e. general and domain factors) before and after adjustment for DIF [66]. The magnitude and impact of DIF effects was used as a basis for assessing the clinical significance of findings [37, 65, 66].

Results

Basic sample characteristics

Basic sample characteristics are summarized in Table 1. The mean age of patients was 38.3 years. Patients were predominantly male and of White ethnicity with a long history of illness. Most patients had a diagnosis of schizophrenia or schizoaffective disorder. Mean CPRS scores indicated that patients were moderately to severely ill. The mean premorbid IQ score of patients was 106.5, and geometric mean completion times of TMT-A and TMT-B were 52.7 and 106.0, respectively.

Table 1 Basic socio-demographic and clinical characteristics at baseline

	Total UK700 sample ($n = 690$)
Age (years), mean (SD)	38.3 (11.6)
Gender, *n* (%)	
Male	395 (57.2)
Female	295 (42.3)
Ethnicity, *n* (%)	
White	361 (52.3)
African Caribbean	190 (27.5)
Other	139 (20.2)
Duration of illness (years)[a], mean (SD)	12.5 (9.7)
Diagnosis, *n* (%)	
Schizophrenia	260 (37.7)
Unspecified functional psychosis	345 (49.1)
Affective disorder	50 (7.3)
Other non-organic psychotic disorder	41 (5.9)
General psychopathology (CPRS)[b], mean (SD)	16.6 (12.4)
Depressive symptoms (MADRS)[c]	10.1 (8.5)
Positive symptoms (CPRS)[c]	4.3 (5.1)
Negative symptoms (SANS)	18.9 (.93)
NART fullscale IQ[d], mean (SD)	106.5 (10.5)
NART verbal IQ[d], mean (SD)	105.4 (11.6)
NART performance IQ[d], mean (SD)	106.9 (8.2)
TMT-A completion times (seconds)[e], geometric mean (SD)	52.7 (26.7)
TMT-B completion times (seconds)[f], geometric mean (SD)	106.0 (50.7)
Treatment group	
Intensive case management	341 (49.4)
Standard case management	349 (50.6)

CPRS comprehensive psychiatric rating scale, *MADRS* Montgomery-Asberg depression rating scale, *SANS* scale for the assessment of negative symptoms, *NART* national adult reading test, *TMT-A* trail making test, part A, *TMT-B* trail making test, part B

Missing values: [a] 2, [b] 34, [c] 1, [d] 108, [e] 94, [f] 255

Dimensional structure of LQOLP

Table 2 displays model fit statistics for unidimensional and bifactor models without and with effects of cognitive deficits and psychiatric symptoms on latent factors and items. A poor model fit was found for the unidimensional model.

Table 2 Model fit statistics for item response models without and with general and specific effects of cognitive deficits and psychiatric symptoms and DIF effects

	Limited information probit (WLSMV) fit statistics				Full information logit (MLR) fit statistics		
	χ^2	CFI	TLI	RMSEA	Log-likelihood	AIC	BIC
Unidimensional model	2,323.71	.455	.728	.172	−22,114.16	44,468.32	45,012.72
Bifactor model[a]	283.59	.957	.978	.049	−21,411.01	43,096.01	43,717.54
Bifactor model[a] with general and specific effects of cognitive deficits (NART, TMT-A, and TMT-B) and general psychopathology (CPRS) adjusted for potential confounders[b]	252.74	.949	.955	.046	−12,208.91	24,907.81	25,882.64
Bifactor model[a] with general and specific effects of cognitive deficits (NART, TMT-A, and TMT-B) and general psychopathology (CPRS) adjusted for potential confounders[b] as well as DIF effects on LQOLP16	249.81	.950	.956	.046	−12,205.93	24,903.86	25,882.66
Bifactor model[a] with general and specific effects of cognitive deficits (NART, TMT-A, and TMT-B) and depressive symptoms (MADRS) adjusted for potential confounders[b]	262.68	.949	.953	.046	−12,583.42	25,656.85	26,639.61
Bifactor model[a] with general and specific effects of cognitive deficits (NART, TMT-A, and TMT-B) and depressive symptoms (MADRS) adjusted for potential confounders[b] as well as DIF effects on LQOLP1 and LQOLP24	254.20	.952	.956	.045	−12,573.83	25,641.65	26,632.44

χ^2 model chi-square, *CFI* comparative fit index, *TLI* Tucker Lewis index, *RMSEA* root mean squared error of approximation, *AIC* Akaike information criterion, *BIC* Bayesian information criterion, *LQOLP1* life as a whole, *LQOLP 16* personal safety, *LQOLP24* mental health, *CPRS* comprehensive psychiatric rating scale, *MADRS* Montgomery-Asberg depression rating scale, *SANS* scale for the assessment of negative symptoms, *NART* national adult reading test, *TMT-A* trail making test, part A, *TMT-B* trail making test, part B, *DIF* differential item functioning

[a] Bifactor model with one general and eight uncorrelated domain factors

[b] Adjusted for: age, gender, ethnicity, length of illness, diagnosis

By comparison, a bifactor model provided a good model fit to the sample data.

Table 3 summarizes factor loadings for the unidimensional and bifactor model. Significant factor loadings on general and domain factors were observed for all LQOLP items in the unidimensional and bifactor model. Comparison of factor loadings across the two models indicated that for most items, factor loadings on the general factor were markedly reduced due to the inclusion of domain factors into the model. In the bifactor model, most items were as discriminating on the general as on the domain factors. Only a few items were markedly more informative on domain factors than the general factor (i.e. 'financial comfort', 'money for enjoyment', 'religious faith', and 'religious practice').

DIF attributable to symptoms and cognitive deficits

Table 4 summarizes findings on DIF attributable to psychiatric symptoms and cognitive deficits. Significant DIF effects were observed for general psychopathology on 'personal safety' ratings as well as depressive symptoms on 'life as a whole' and 'mental health' ratings. Patients with the same SQOL but higher psychopathology scores were less likely to rate their 'personal safety' positively. The magnitude of the proportional odds ratio for the DIF effect of general psychopathology on 'personal safety' ratings was, however, small (OR .96, 95% CI .93–.99). Further, the impact of DIF was found to be limited when comparing associations of general psychopathology with general and domain factors before (G, general SQOL factor, B −.017, 95% CI −.022 to −.010; D5, safety, B −.005, 95% CI −.013 to .003) and after (G, general SQOL factor, B −.016, 95% CI −.023 to −.010; D5, safety, B .007, 95% CI −.004 to .018) adjustment for DIF. Similarly, patients with higher depressive symptom scores were less likely to endorse more positive 'life as a whole' and 'mental health' ratings. Again, the proportional odds ratios for the DIF effects were of small magnitude ('life as a whole', OR .93, 95% CI .89 to .98; 'mental health', OR .93, 95% CI .91 to .97). Moreover, the impact of the DIF effects was small, as the strength of the association of depressive symptoms with the general SQOL factor and health domain factor was only slightly attenuated before (G, general SQOL factor, B −.030, 95% CI −.038 to −.022; D8, health, B −.009, 95% CI −.017 to −.002) and after (G, general SQOL factor, B −.020, 95% CI −.033 to −.007; D8, health, B −.004, 95% CI −.015 to .007) adjustment for DIF.

Table 3 Factor loadings (standard errors) in bifactor model of LQOLP ratings

Items		Unidimensional model	Bifactor model							
		G	G	D1	D2	D3	D4	D5	D6	D7	D8
LQOLP1	Life as a whole	.46 (.03)***	.54 (.04)***								
LQOLP2	Job situation	.32 (.04)***	.36 (.04)***								
LQOLP3	Financial comfort	.57 (.02)***	.28 (.04)***	.84 (.02)***							
LQOLP4	Money for enjoyment	.59 (.02)***	.32 (.04)***	.84 (.02)***							
LQOLP5	Getting on with people	.54 (.03)***	.57 (.03)***		.54 (.03)***						
LQOLP6	Number of friends	.53 (.03)***	.55 (.03)***		.54 (.03)***						
LQOLP7	Pleasure home acts	.55 (.03)***	.56 (.03)***			.57 (.05)***					
LQOLP8	Pleasure outside acts	.50 (.03)***	.50 (.03)***			.49 (.05)***					
LQOLP9	Pleasure radio/TV	.43 (.03)***	.43 (.04)***			.41 (.05)***					
LQOLP10	Living arrangements	.67 (.02)***	.54 (.03)***				.48 (.03)***				
LQOLP11	Residence privacy	.62 (.02)***	.46 (.04)***				.55 (.04)***				
LQOLP12	Continued residence	.61 (.02)***	.51 (.04)***				.56 (.03)***				
LQOLP13	Residence independence	.74 (.02)***	.48 (.04)***				.72 (.03)***				
LQOLP14	Residence influence	.75 (.02)***	.54 (.04)***				.63 (.03)***				
LQOLP15	Other residents	.63 (.02)***	.53 (.04)***				.44 (.04)***				
LQOLP16	Personal safety	.61 (.02)***	.53(.03)***					.67 (.02)***			
LQOLP17	Neighbourhood safety	.61 (.02)***	.52 (.03)***					.67 (.02)***			
LQOLP18	Family situation	.44 (.03)***	.44 (.04)***						.63 (.03)***		
LQOLP19	Amount family contact	.41 (.03)***	.41 (.04)***						.63 (.03)***		
LQOLP20	Religious faith	.20 (.04)***	.23 (.05)***							.52 (.04)***	
LQOLP21	Religious practice	.16 (.04)***	.18 (.05)***							.52 (.04)***	
LQOLP22	Frequency doctor	.46 (.03)***	.49 (.03)***								.31 (.05)***
LQOLP23	General health	.51 (.03)***	.53 (.03)***								.61 (.08)***
LQOLP24	Mental health	.46 (.03)***	.48 (.04)***								.42 (.06)***

Model 1 Unidimensional model, *Model 2* Bifactor model, *G* general SQOL factor, *D* domain factor, *D1* finances, *D2* social relations, *D3* leisure, *D4* living situation, *D5* safety, *D6* family relations, *D7* religion, *D8* health

Model estimation: limited information probit (WLSMV)

* $P < .05$; ** $P < .01$; *** $P < .001$

Table 4 Estimates (standard errors) of DIF effects of cognitive deficits and psychiatric symptoms on LQOLP ratings in adjusted bifactor model

Covariates	DIF effects		
	LQOLP1 Life as a whole	LQOLP16 Personal safety	LQOLP24 Mental health
General psychopathology (CPRS)	–	−.016 (.004)***	–
Depressive symptoms (MADRS)	−.038 (.009)***	–	−.034 (.008)***
Positive symptoms (CPRS)	–	–	–
Negative symptoms (SANS)	–	–	–
NART fullscale IQ	–	–	–
NART verbal IQ	–	–	–
NART performance IQ	–	–	–
Trail A completion times	–	–	–
Trail B completion times	–	–	–

CPRS comprehensive psychiatric rating scale, *MADRS* Montgomery-Asberg depression rating scale, *SANS* scale for the assessment of negative symptoms, *NART* national adult reading test, *TMT-A* trail making test, part A, *TMT-B* trail making test, part B, *SE* standard error, *DIF* differential item functioning

[a] Adjusted for general and specific effects of: age, gender, ethnicity, length of illness, diagnosis

[b] Model estimation: limited information probit (WLSMV)

*** $P < .001$

Discussion

Main findings

This study examined whether LQOLP ratings are affected by DIF attributable to psychiatric symptoms and cognitive deficits in a large sample of patients with severe and enduring psychosis. We found evidence of DIF for 'personal safety', 'life as a whole', and 'mental health' ratings, such that the probability of item responses was altered for patients with the same SQOL but with different levels of symptom severity. Yet, the magnitude and impact of DIF effects were small. There was no evidence that LQOLP ratings were biased by cognitive deficits.

Methodological considerations

Patients included in the study were recruited in the context of a RCT and not randomly selected. Findings may thus not be readily generalizable to patients in routine care. Some caution must be taken in interpreting the results due to missing data for measures of cognitive deficits. It cannot be ruled out that those patients who did not complete the cognitive tests were those most adversely affected by their illness. Further, the cognitive tests used in the current study employ stimuli with relatively low relevance to tasks performed in ordinary life [67]. This may have limited the ecological validity of findings.

DIF was examined using the approach of MIMIC models. This method can detect non-uniform DIF (as uniform DIF) only if the DIF results in a shift in the conditional probability of an indicated response. We still considered MIMIC models as the most appropriate method as in the absence of a priori hypothesis on the type of DIF items are expected to have a safer inference is to determine any type of DIF [37]. There is also evidence from simulation studies on inflated type-I error associated with the assumption of the MIMIC approach that if one item is tested for DIF, all other items are DIF free [50]. However, Woods [50] noted that this is not particularly problematic if each item including those assumed to be DIF free is subsequently tested for DIF. Moreover, MIMIC models offer several important advantages compared to other approaches of DIF detection such as less restrictive sample size requirements, greater accuracy with small focal-group samples, and control of multiple covariates, which may be categorical or continuous [50].

Comparisons with previous research

Many studies have investigated psychiatric symptoms and cognitive deficits as factors associated with SQOL ratings [18, 20]. Correlations of symptoms and deficits with SQOL sum scores provide, however, only limited evidence on whether these factors bias items and impair the validity of self-report measures. The current study has moved beyond previous research by distinguishing effects of factors that bias patients' responses to specific items from those influencing their underlying SQOL [28].

There was evidence that responses to some LQOLP items are biased by symptom severity. Specifically, we found that the 'personal safety' item of the LQOLP was

affected by DIF attributable to general psychopathology. A higher prevalence of violent victimization has been reported for psychotic patients with more severe symptoms [68]. This might explain why patients with more severe psychopathology but with the same SQOL were more likely to have lower ratings of their satisfaction with 'personal safety'. Similarly, patients with more severe depressive symptoms were more likely to rate their 'life as a whole' and 'mental health' negatively. This is consistent with the previously published literature, in that ratings of more general items have been reported to be more prone to affective bias [69, 70].

Taken together, findings suggest that only three out of twenty-four LQOLP items were affected by DIF due to psychiatric symptoms. This is at variance with earlier concerns that patient ratings of SQOL may be consistently biased by psychiatric symptoms [18, 20]. In contrast to previous studies [13, 14], we found no evidence to suggest that cognitive deficits would differentially affect LQOLP ratings.

Implications

Structural validity is a central property for determining the value of established outcome measures [71]. Our findings suggest that the structural validity of a widely used measure of SQOL, i.e., the LQOLP, may be, to a degree, affected by DIF attributable to psychiatric symptoms. In other words, scores may to some extent capture concepts (i.e. psychiatric symptoms) other than the one the LQOLP is intended to measure. However, DIF was present in only a small proportion of items, and the magnitude and impact of DIF we found for the LQOLP is unlikely of clinical significance [65].

In the light of growing interest in elucidating whether health outcome measures may function inappropriately or unfairly and thereby contribute to ongoing health disparities [37, 72], psychosis research faces the challenge to enhance the validity of established PROs by addressing measurement issues such as DIF. There are several ways of addressing this. Biased items may be recommended for expert review and, ultimately, exclusion [29]. However, one may not want to exclude items if they are of particular conceptual importance. For instance, it has been suggested that psychotic patients' satisfaction with their mental health is a central aspect of their SQOL [1, 63]. Satisfaction with mental health, although subject to DIF, may therefore be seen as an essential item. In such cases, one can keep the item in the instrument and adjust for DIF statistically. However, statistical adjustment for DIF requires model-based approaches to scoring such as item response modelling and item banking, which have not yet been developed for SQOL in patients with psychosis [6, 73].

In summary, the study adds new evidence to the debate on the validity of SQOL measures in psychotic patients. Testing the impact of symptom levels and cognitive deficits on SQOL item responses, there is little evidence to suggest that DIF attributable to symptoms is of clinical significance and no evidence of DIF due to deficits. On balance this may rather encourage the use of SQOL measures in patients with psychosis.

Acknowledgments This work was supported by a Research Training Fellowship funded by the National Institute of Health Research, UK, to U.R. The report is independent research and the views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

References

1. Priebe, S., & Fakhoury, W. (2008). Quality of life. In K. Mueser & D. Jeste (Eds.), *The clinical handbook of schizophrenia* (pp. 581–591). New York: Guildford.
2. Lehman, A. (1996). Measures of quality of life among persons with severe and persistent mental disorders. *Social Psychiatry and Psychiatric Epidemiology, 31*, 78–88.
3. Food, U. S., & Drug Administration (FDA). (2009). *Guidance for industry. Patient-reported outcome measures: Use in medical product development to support labelling claims*. Rockville, MD: US FDA.
4. Institute of Medicine, Committee on Quality of Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.
5. Prieto, L., Novick, D., Sacristan, J., Edgell, E., Alonso, J., & on behalf of the SOHO Study Group. (2003). A Rasch model analysis to test the cross-cultural validity of the EuroQoL-5D in the schizophrenia outpatient health outcomes study. *Acta Psychiatrica Scand, 107*, 24–29.
6. Reininghaus, U., McCabe, R., Burns, T., Croudace, T., & Priebe, S. (2010). Measuring patients' views: A bi-factor model of distinct patient-reported outcomes in psychosis. Psychological Medicine. doi:10.1017/S0033291710000784.
7. UK Department of Health. (2008). *High quality care for all. NHS next stage review final report*. London, UK: Department of Health.
8. Sponheim, S., Jung, R., Seidman, L., Mesholam-Gately, R., Manoach, D., O'Leary, R., et al. (2010). Cognitive deficits in recent-onset and chronic schizophrenia. *Journal of Psychiatric Research, 44*, 421–428.
9. Braga, R., Mendlowicz, M., Marrocos, R., & Figueira, I. (2005). Anxiety disorders in outpatients with schizophrenia: Prevalence and impact on the subjective quality of life. *Journal of Psychiatric Research, 39*, 409–414.
10. Cuesta, M., Peralta, V., Gil, P., & Artamendi, M. (2003). Psychopathological dimensions in first-episode psychoses. From the trunk to the branches and leaves. *European Archives of Psychiatry and Clinical Neuroscience, 253*, 73–79.
11. Demjaha, A., Morgan, K., Morgan, C., Landau, S., Dean, K., Reichenberg, A., et al. (2009). Combining dimensional and categorical representation of psychosis: The way forward for DSM-V and ICD-11? *Psychological Medicine, 39*, 1943–1955.
12. Smith, D., Mar, C., & Turoff, B. (1998). The structure of schizophrenic symptoms: A meta-analytic confirmatory factor analysis. *Schizophrenia Research, 31*, 57–70.

13. Addington, J., & Addington, D. (2000). Neurocognitive and social functioning in schizophrenia: A 2.5 year follow-up study. *Schizophrenia Research, 44*, 47–56.
14. Corrigan, P., & Buican, B. (1995). The construct validity of subjective quality of life for the severely mentally ill. *Journal of Nervous and Mental Disease, 183*, 281–285.
15. Green, M., Kern, R., & Heaton, R. (2004). Longitudinal studies of cognition and functional outcome in schizophrenia: Implications for MATRICS. *Schizophrenia Research, 72*, 41–51.
16. Eack, S., & Newhill, C. (2007). Psychiatric symptoms and quality of life in schizophrenia. *Schizophrenia Bulletin, 33*, 1225–1237.
17. Priebe, S., Reininghaus, U., McCabe, R., Burns, T., Eklund, M., Hansson, L., et al. (2010). Factors influencing subjective quality of life in patients with schizophrenia and other mental disorders: A pooled analysis. Schizophrenia Research. doi:10.1016/j.schres.2009.12.020.
18. Atkinson, M., Zibin, S., & Chuang, H. (1997). Characterizing quality of life among patients with chronic mental illness: A critical examination of the self-report methodology. *American Journal of Psychiatry, 154*, 99–105.
19. Holloway, F., & Carson, J. (2002). Quality of life in severe mental illness. *International Review of Psychiatry, 14*, 175–184.
20. Katschnig, H. (1997). How useful is the concept of quality of life in psychiatry? *Current Opinion in Psychiatry, 10*, 337–345.
21. Pavot, W., & Deiner, E. (1993). The affective and cognitive context of self-reported measures of subjective well-being. *Social Indicators Research, 28*, 1–20.
22. Epstein, A., Hall, J., Tognetti, J., Son, L., & Conant, L. (1989). Using proxies to evaluate quality of life: Can they provide valid information about patients' health status and satisfaction with medical care? *Medical Care, 27*, s91–s98.
23. Jenkins, C. (1992). Assessment of outcomes of health intervention. *Social Science and Medicine, 35*, 367–375.
24. Rosenblatt, A., & Attkisson, C. (1993). Assessing outcomes for sufferers of severe mental disorder: A conceptual framework and review. *Evaluation and Program Planning, 16*, 347–763.
25. Corring, D. (2002). Quality of life: Perspectives of people with mental illnesses and family members. *Psychiatric Rehabilitation Journal, 25*, 350–358.
26. Herrman, H., Hawthorne, G., & Thomas, R. (2002). Quality of life assessment in people living with psychosis. *Social Psychiatry and Psychiatric Epidemiology, 37*, 510–518.
27. Lasalvia, A., Ruggeri, M., & Santolini, N. (2002). Subjective quality of life: Its relationship with clinician-rated and patient-rated psychopathology—The South-Verona Outcome project 6. *Psychotherapy and Psychosomatics, 71*, 275–284.
28. Mellenbergh, G. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127–143.
29. Teresi, J., Ocepek-Welikson, K., Kleinman, M., Eimicke, J., Crane, P., Jones, R., et al. (2009). Analysis of differential item functioning in the depression item bank from the patient reported outcome measurement information system (PROMIS): An item response theory approach. *Psychological Science Quarterly, 51*, 148–180.
30. Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Hillsdale, NJ: Lawrence Erlbaum Associates.
31. Oliver, J., Huxley, P., Priebe, S., & Kaiser, W. (1997). Measuring the quality of life of severely mentally ill people using the Lancashire quality of life profile. *Social Psychiatry and Psychiatric Epidemiology, 32*, 76–83.
32. Van Nieuwenhuizen, C., Schene, A., Koeter, M., & Huxley, P. (2001). The Lancashire quality of life profile: Modification and psychometric evaluation. *Social Psychiatry and Psychiatric Epidemiology, 36*, 36–44.
33. Gaite, L., Vazquez-Barquero, J., Arrizabalaga, A., Schene, A., Welcher, B., Thornicroft, G., et al. (2000). Quality of life in schizophrenia: Development, reliability and internal consistency of the Lancashire quality of life profile—European Version. *British Journal of Psychiatry, 177*, s49–s54.
34. Ruggeri, M., Warner, R., Bisoffi, G., & Fontecedro, L. (2001). Subjective and objective dimensions of quality of life in psychiatric patients: A factor analytic approach. *British Journal of Psychiatry, 178*, 268–275.
35. Rose, M., Bjorner, J., Becker, J., Fries, J., & Ware, J. (2008). Evaluation of a preliminary physical function item bank supported the expected advantages of the patient-reported outcomes measurement information system (PROMIS). *Journal of Clinical Epidemiology, 61*, 17–33.
36. Reeve, B., Hays, R., Chang, C.-H., & Perfetto, E. (2007). Applying item response theory to enhance health outcomes assessment. *Quality of Life Research, 16*, 1–3.
37. Yang, F., Tommet, D., & Jones, R. (2009). Disparities in self-reported geriatric depressive symptoms due to socio-demographic differences: An extension of the bi-factor item response theory model for use in differential item functioning. *Journal of Psychiatric Research, 43*, 1025–1035.
38. Burns, T., Creed, F., Fahy, T., Thompson, S., Tyrer, P., White, I., et al. (1999). Intensive versus standard case management for severe psychotic illness: A randomised trial. *Lancet, 353*, 2185–2189.
39. Spitzer, R., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry, 35*, 773–782.
40. Asberg, M., Montgomery, S., Perris, C., Schalling, D., & Sedvall, G. (1978). A comprehensive psychopathological rating scale. *Acta Psychiatrica Scandinavica, 271*, 5–27.
41. Montgomery, S., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry, 134*, 382–389.
42. Andreasen, N. (1989). Scale for the assessment of negative symptoms (SANS). *British Journal of Psychiatry, 7*, 53–58.
43. Nelson, H., & Williams, J. (1991). *The revised national adult reading test manual*. Slough, UK: National Foundation for Educational Research.
44. Crawford, J., Stewart, L., Cochrance, R., Foulds, J., Besson, J., & Parker, D. (1989). Estimating premorbid IQ from demographic variables: Regression equations derived from a UK sample. *British Journal of Clinical Psychology, 28*, 275–278.
45. Reitan, R. (1958). Validity of the trail making test as an indicator of organic brain dysfunction. *Perceptual and Motor Skills, 8*, 271–276.
46. Lezak, M. (1983). *Neuropsychological assessment*. New York: Oxford University Press.
47. Muthén, L., & Muthén, B. (1998–2010). Mplus version 5.2. Los Angeles: Muthén and Muthén.
48. Joereskog, K., & Goldberger, A. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of American Statistical Association, 70*, 631–639.
49. Muthén, B. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. *Journal of Educational Statistics, 10*, 121–132.
50. Woods, C. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research, 44*, 1–27.
51. Muthén, B. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In

H. Wainer & H. Braun (Eds.), *Test validity* (pp. 213–238). Hillsdale, NJ: Lawrence Erlbaum.
52. Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Meetings of psychometric society (Los Angeles, California and Leuven, Belgium). *Psychometrika, 54*, 557–585.
53. Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.
54. Steiger, J. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.
55. Tucker, L., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10.
56. Akaike, H. (1987). Prediction and entropy. In A. Atkinson & S. Fienberg (Eds.), *A celebration of statistics* (pp. 1–24). New York: Springer.
57. Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
58. Gibbons, R., & Hedeker, D. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*, 423–436.
59. Gibbons, R., Bock, D., Hedeker, D., Weiss, D., Segawa, E., Bhaumik, D., et al. (2007). Full-information bifactor analysis for graded response data. *Applied Psychological Measurement, 31*, 4–19.
60. Gibbons, R., Weiss, D., Kupfer, D., Frank, E., Fagioloini, A., Grochocinski, V., et al. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services, 59*, 361–368.
61. Gibbons, R., Rush, A., & Immekus, J. (2009). On the psychometric validity of the domains of the PDSQ: An illustration of the bi-factor item response theory model. *Journal of Psychiatric Research, 43*, 401–410.
62. Reise, S., Morizot, J., & Hays, R. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19–31.
63. Lehman, A. (1988). A quality of life interview for the chronically mentally ill. *Evaluation Program Planning, 11*, 51–62.
64. Vatne, S., & Bjorkly, S. (2008). Empirical evidence for using subjective quality of life as an outcome variable in clinical studies: A meta-analysis of correlates and predictors in persons with a major mental disorder living in the community. *Clinical Psychology Review, 28*, 869–889.
65. Cole, S., Kawachi, I., Maller, S., & Berkman, J. (2000). Test of item-response bias in the CES-D scale: Experience from the New Haven EPESE study. *Journal of Clinical Epidemiology, 53*, 285–289.
66. Teresi, J. (2006). Different approaches to differential item functioning in health applications: Advantages, disadvantages and some neglected topics. *Medical Care, 44*, s152–s170.
67. Bentall, R. (2003). *Madness explained. Psychosis and nature*. London: Penguin Books.
68. Walsh, E., Moran, P., Scott, C., McKenzie, K., Burns, T., Creed, F., et al. (2003). Prevalence of violent victimisation in severe mental illness. *British Journal of Psychiatry, 183*, 233–238.
69. Stone, A., & Litcher-Kelly, L. (2006). Momentary capture of real world data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement* (pp. 61–72). Washington, DC: American Psychological Association.
70. Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahnemann, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). New York: Russell Sage Foundation.
71. Mokkink, L., Terwee, C., Knol, D., Stratford, P., Alonso, J., Patrick, D., et al. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology, 10*(22). doi:10.1186/1471-2288-10-22
72. US Department of Health, Human Services. (2000). *Healthy people 2010: Understanding and improving health*. Washington, DC: US Government Printing Office.
73. Fries, J., Bruce, B., & Cella, D. (2005). The promise of PROMIS: Using item response theory to improve assessment of patient-reported outcomes. *Clinical Experimental Rheumatology, 23*, s53–s57.