npg

Development of an Abbreviated Schizophrenia Quality of Life Scale Using a New Method

Warren B Bilker[*,1,2], **Colleen Brensinger**[1,2], **Matthew M Kurtz**[1], **Christian Kohler**[1], **Ruben C Gur**[1], **Steven J Siegel**[1] and **Raquel E Gur**[1]

[1]*Schizophrenia Research Center, Neuropsychiatry Section, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA;* [2]*Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*

The goal of the study was to develop and apply a predictive model approach to reduce the number of items collected for scales that yield a total summary score. A parsimonious subset of items from the 21-item Quality of Life Scale (QLS) that can accurately predict the total scale score was sought and evaluated in 198 patients with schizophrenia, using a statistical modeling approach. Two additional data sets were used for model validation: the subset of 101 patients used in the model construction who had the QLS administered approximately 1 year later and a new sample of 37 patients. Using only seven QLS items as predictors, the correlation was 0.9831 between the predicted and true QLS totals. Applying the model based for these seven QLS items, the correlations from the first and second validation data sets were 0.9791 and 0.9637, respectively. The study demonstrates that a small subset of items of the QLS predicts the entire 21-item scale with high accuracy. Two validation samples have confirmed the finding. This reduces the effort associated with scale administration and is likely to increase the assessment of an important functional domain. Such models can guide efforts for item reduction in other rating instruments.
Neuropsychopharmacology (2003) **28,** 773–777. doi:10.1038/sj.npp.1300093

Keywords: quality of life scale; scale reduction; predictive model

INTRODUCTION

Standardized assessment of symptoms and functioning requires a lengthy battery of clinical scales. A systematic approach is needed to reduce the number of items collected for each assessment scale.

Traditional methods for scale development include item analysis, which aims at weeding out items that are insufficiently correlated with the scale total. However, a more rigorous procedure is needed when the goal is to identify a subset of items that accurately predict the totals for established clinical scales. We present a new approach to reduce the number of items collected for existing scales that yield a total summary score. The method seeks a parsimonious subset of scale items that, when assessed in place of the full set of scale items, can predict the scale total with good accuracy. The approach is a variant of the all subsets regression approach (Hocking and Leslie, 1967; Draper and Smith, 1998), where the criteria used to rank the

*Correspondence: Dr W Bilker, Neuropsychiatry Section, University of Pennsylvania Medical Center, 10th Floor Gates Bldg, Philadelphia, PA 19104, USA, Tel: +1 215 662 6095, Fax: +1 215 662 7093, E-mail: wbilker@cceb.upenn.edu

models is the correlation between the actual scale total and the model predicted scale total, rather than the model R^2. Model validation, using a secondary source of data, is critical in this approach.

We examined the Quality of Life Scale (QLS; Henrichs *et al*, 1984), a 21-item scale commonly used as a measure of functioning in schizophrenia, to illustrate the use of this new approach to scale reduction. The QLS balances subjective questions regarding life satisfaction and objective indicators of social and occupational role functioning. When administered by a trained clinician as a semi-structured interview, the scale provides information on symptoms and functioning during the 4 weeks prior to assessment. Behavioral anchors are presented for each item, scored on a 0 (severe impairment) to 6 (high functioning) scale. The QLS assesses four interdependent theoretical constructs: intrapsychic foundations (items 13–17, 20, 21), consisting of measures related to sense of purpose and motivation; interpersonal relations (items 1–8), examining social experience; instrumental role (items 9–12), related to work functioning; and common objects and activities (items 18–19), which measure engagement in the community by possession of common objects and participation in a range of activities (Henrichs *et al*, 1984). These theoretical distinctions were supported by principal components factor analysis conducted on data collected from 111 patients with schizophrenia.

The utility of the QLS in assessing change in deficit status has been highlighted in treatment studies (eg Meltzer et al, 1990; Rosenheck et al, 1997; Hamilton et al, 1998). The importance of the QLS in assessing outcome in schizophrenia research has also been highlighted by the observed relation between neurocognitive impairment and psychosocial status (Green, 1996; Buchanan et al, 1994; Medalia et al, 1998; Wykes et al, 1999).

The QLS requires 30–45 min of administration time by a trained clinician. In addition, since there is overlap between several items on the QLS and other scales, participants' engagement and cooperation with the interview can wane. Several studies have suggested that a subset of individual items on the QLS are highly related, and may provide an index of performance on the scale as a whole. For example, in their original formulation and description of the QLS, Henrichs et al (1984) showed that 52% of the variance in their study was explained by an Interpersonal Relations factor consisting of items 1–8 of the scale, while only 22% of the variance was explained by the other three factors combined. More recently, in a study of sex differences in clinical expression of schizophrenia (Shtasel et al, 1992), we showed that a factor analysis of QLS ratings for 107 patients with schizophrenia revealed three factors: (1) social functioning (items 2–8, 16), (2) engagement (items 1, 13–15, 18–21), and (3) vocational functioning (items 9–11, 12, 17). Others have noted that the four subscales of the QLS are highly related, and provide little additional information beyond the total score. Meltzer et al (1990) reported that both before and after clozapine treatment, all QLS scores from their patient sample were highly intercorrelated and correlated very highly with the total QLS score. Recent investigations of the relation between symptom exacerbation and stabilization and Quality of Life (QOL) have reported QLS total scores, rather than subscale scores, in the light of each subscale's very close relation with the total score (Bow-Thomas et al, 1999). The goal of the present study was to identify the smallest subset of items from the original QLS scale that can predict the total QLS score.

METHOD

Participants

Patients with schizophrenia were consecutive recruits for brain–behavior studies at the University of Pennsylvania Schizophrenia Center. After informed consent was obtained, participants underwent standard comprehensive screening , including medical history, physical examination, laboratory tests, and assessment procedures (Shtasel et al, 1992; Gur et al, 1991). These included the patient edition of the SCID (SCID-P; Spitzer et al, 1996), and scales for measuring symptomatology and functioning administered by investigators trained to a criterion reliability of 0.90, intraclass correlation (Shtasel et al, 1992). Entry criteria included (a) a diagnosis of schizophrenia or schizophreniform disorder by DSM-IV criteria, (b) no concomitant axis I or II disorder, including past or present substance abuse or dependence, and (c) no medical or neurological disorder that may affect brain function. There were 198 patients satisfying these criteria who had the 21-item QLS administered at study entry based on reliable information obtained from patients, family, and care providers. Two validation data sets were used: (1) the subset of 101 patients who had the QLS administered approximately 1 year after study entry, and (2) 37 patients recruited subsequently and satisfying the entry criteria. Demographic and clinical information for the three groups of patients is provided in Table 1.

Data Analysis

The goal was to determine a parsimonious subset of the 21 items in the QLS that can be used to predict the total of the 21 items with 'optimal' prediction properties. To determine an optimal combination of items, all combinations of QLS items were considered for all subsets containing 1–10 items (a total of 1 048 575 models). The maximum of 10 items considered represent a reduction to less than one half of the 21 scale items collected for the full-scale administration. The approach is similar to all subsets regression, except that the model performance statistic assessed for each combination of items was the Pearson correlation coefficient, ρ, between the total QLS and the predicted total QLS, rather than the model R^2.

Two data sets were used for model validation. The first was a subset of 101 of the 198 patients used for the model construction, who had the QLS administered approximately 1 year after study entry. The second was comprised of 37 participants recruited during the period when this analysis was underway. For each subject in each validation data set, the best model for each number of predictors (1–10 QLS

Table 1 Characteristics of Samples

Characteristic	Model construction sample (n = 198)	Follow-up sample (n = 101)	New sample (n = 37)
Gender (% male)	61.1	62.4	59.5
Race (% white)	52.5	48.5	43.2
Medication naïve (%)	31.3	41.6	38.2
Patient first episode (%)	32.3	40.6	51.4
Mean ± SD (range)			
Age (years)	32.7 ± 10.9 (17–73)	32.4 ± 8.8 (19–63)	33.0 ± 11.0 (16–64)
Age at onset (years)	23.6 ± 7.1 (11–49)	23.2 ± 6.8 (13–48)	26.1 ± 7.7 (15–46)
Duration of illness (years)	8.9 ± 8.9 (0–42)	9.3 ± 7.0 (1–34)	7.7 ± 10.3 (0–39)
Education (years)	12.8 ± 2.2 (6–22)	13.0 ± 2.2 (8–22)	13.6 ± 2.7 (9–20)
Total QLS score	50.4 ± 27.1 (0–121)	59.6 ± 28.8 (6–122)	66.7 ± 25.4 (14–115)

items) was used to predict the total QLS. For each validation data set, the correlation was assessed for each model.

RESULTS

The largest correlation between the total QLS and the predicted total QLS for models having 1–10 items as predictors are presented in Table 2 (under 'model construction sample'), while the associated model coefficients are presented in Table 3. With a single item, the largest correlation achieved for predicting the total QLS is 0.8689, which is for item 14. The correlation is 0.8234 using only item 17, 0.7643 using only item 10, and remains at or above 0.7500 for items 7, 6, and 11. Using two predictors, the highest correlation achieved was 0.9224, which is for a model with items 7 and 14. From Table 3, the best prediction model for two QLS predictors is predicted total QLS = 11.278+6.285*QLS7+9.627*QLS14. It is important to point out that there are seven other combinations of QLS items that yield correlations above 0.9100. These are models containing items 6 and 14 ($\rho = 0.9196$), 3 and 14 ($\rho = 0.9191$), 4 and 14 ($\rho = 0.9142$), 4 and 10 ($\rho = 0.9141$), 6 and 17 ($\rho = 0.9137$), 14 and 17 ($\rho = 0.9135$), and 2 and 14 ($\rho = 0.9104$).

Several approaches were applied to determine the 'optimal' number of items considering the tradeoff between parsimony and correlation. An informal approach was to examine the relative gains in correlations for increasing numbers of predictors. The gain in correlation from adding predictors decreases for a greater number of predictors, as illustrated in Figure 1. There is a 0.58% gain in correlation for the top models increasing from 5 to 6 item predictors, 0.52% increasing from 6 to 7 predictors, 0.38% moving from 7 to 8, 0.27% moving from 8 to 9, and 0.13% moving from 9 to 10 predictors. Multiple formal testing approaches were also applied. These included bootstrapping and permutation methods for testing the statistical significance of the increases in correlation of adding each predictor as well as a paired t-test to test the decrease in absolute error associated with each additional predictor. Each of these formal statistical testing approaches found statistical significance even in the miniscule increase in correlation of 0.0013 in moving from 9 to 10 predictors. Thus, the more formal approaches proved not to be helpful because of their sensitivity beyond what is clinically significant. However, applying an *a priori* rule to add predictors until the gain in correlation fell below 0.5%, the 'optimal' model includes seven predictors. A plot of the actual total QLS *vs* the predicted total QLS for these data for seven predictors is shown in Figure 2.
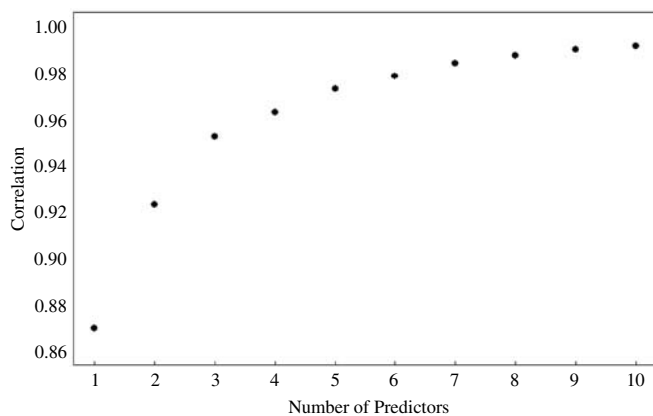
The items included in this top model with seven predictors are 3, 6, 9, 14, 16, 18, and 20. From Table 3, the 'optimal' prediction model for seven predictors is

Table 2 Results for Top Models with 1–10 QLS Items from Original Sample and from Validation Data Sets

No. of predictors	Model construction sample (n = 198)		Follow-up sample (n = 101)		New sample (n = 37)	
	ρ	Mean ± SD	ρ	Mean ± SD	ρ	Mean ± SD
1	0.8689	50.4 ± 23.5	0.8115	59.7 ± 25.3	0.8091	70.5 ± 23.8
2	0.9224	50.4 ± 25.0	0.8776	61.3 ± 27.2	0.9258	67.7 ± 24.3
3	0.9515	50.4 ± 25.8	0.9264	60.4 ± 27.9	0.9323	70.6 ± 24.3
4	0.9623	50.4 ± 26.1	0.9460	60.4 ± 28.2	0.9386	71.5 ± 24.7
5	0.9724	50.4 ± 26.3	0.9654	60.1 ± 28.0	0.9532	71.7 ± 24.9
6	0.9780	50.4 ± 26.5	0.9700	60.4 ± 28.5	0.9627	71.3 ± 24.9
7	0.9831	50.4 ± 26.6	0.9791	59.6 ± 28.0	0.9637	70.9 ± 24.9
8	0.9868	50.4 ± 26.7	0.9786	59.9 ± 28.2	0.9670	71.2 ± 24.5
9	0.9895	50.4 ± 26.8	0.9831	59.7 ± 28.1	0.9769	70.9 ± 24.0
10	0.9908	50.4 ± 26.8	0.9844	59.7 ± 28.0	0.9767	70.5 ± 24.2

Table 3 Coefficients for Top Models with 1–10 Predictors (n = 198)

No. of predictors	QLS1	QLS3	QLS5	QLS6	QLS7	QLS9	QLS14	QLS16	QLS17	QLS18	QLS20	Intercept
1							12.989					18.937
2					6.285		9.627					11.278
3			5.230				6.213		4.569			15.802
4			5.314				4.879		4.083		2.972	10.792
5			5.487			2.367	3.716		2.922		3.247	10.858
6		2.252	3.912			2.346	3.753		2.711		3.100	10.325
7		2.253	3.530			2.770	4.031	2.110		2.233	2.452	3.797
8		2.114	3.495			2.246	3.435	1.755	1.853	1.880	2.288	5.030
9		1.801	1.625	2.882		2.314	3.144	1.748	1.966	1.813	2.184	2.901
10	−1.106	1.756	1.326	2.801		2.266	3.063	1.652	1.994	1.774	1.963	2.247

Figure 1 Plot of number of predictors by correlation for best model.

Figure 2 Plot of total QLS by predicted total QLS for seven predictors for the model construction data set.

total QLS $= 3.797+2.253*QLS3+3.530*QLS6+2.770*QLS9+4.031*QLS14+2.110*QLS16+2.233*QLS18+2.452*QLS20$. There are three other seven-item models that also exceed a correlation of 0.9820. All three models contain items 6, 9, 14, 20, and at least two of items 16, 17, or 18. The second ranked model includes items 3, 6, 9, 14, 17, 18, and 20, which replaces item 16 with 17 from the top ranked model, and has a correlation of 0.9829.

It is important to validate the predictive ability of the models with different data. The highest correlation achieved for seven predictors is 0.9831. Two data sets were used to validate the top models for each number of predictors: the subset of patients used in the model construction who had the QLS administered approximately 1 year after study entry, and a new sample of 37 patients not included in the model construction data set. The top models for each number of QLS predictors, as presented in Table 3, were applied to the subject's data in the validation data sets. The correlations for each of these validated top models are presented in Table 2. The correlations for the validation data sets are minimally lower than those from the data used to construct the models. The original correlation from the seven-predictor model was 0.9831, while the correlations from the follow-up and new sample validation data sets are 0.9791 and 0.9637, respectively. The relatively small drop in correlation for the two validation data sets holds for all numbers of predictors considered, 1–10. In the new sample validation data set, there is a +1.00% change in correlation moving from 5 to 6 items, +0.10% from 6 to 7, +0.34% from 7 to 8, +1.02% from 8 to 9, and −0.02% moving from 9 to 10 predictors. In the follow-up validation data set, the changes were +0.48% moving from 5 to 6 predictors, +0.94% from 6 to 7, −0.05% from 7 to 8, +0.46% moving from 8 to 9, and +0.13% moving from 9 to 10 predictors.

For all numbers of items considered, the ordering of the models for all possible combinations of items, as well as the formal and informal decision rule outcomes on the number of 'optimal' predictors, was exactly the same if ordered by the intraclass correlation coefficient (Shrout and Fleiss, 1979) or the mean squared error (MSE) for the models, rather than the correlation. These results are all consistent with and validate the choice of seven predictors.
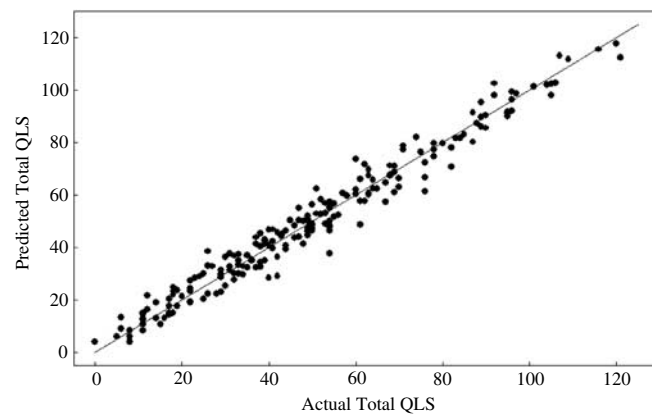
To examine the effects of ethnicity we divided the sample into Caucasians and non-Caucasians, and obtained the correlations for each race group using the top model for each number of predictors in Table 3, which were based on the whole original sample. The correlations differed negligibly.

DISCUSSION

The study has demonstrated that seven items of the QLS can provide an excellent prediction of the entire 21-item QLS. Two validation samples have confirmed the finding. Thus, an abbreviated scale can substitute the full scale with superior accuracy. This reduces considerably the effort associated with the assessment and is likely to result in enhancing the utility of the scale. The abbreviated QLS includes items representing all four interdependent theoretical constructs of the original 21-item QLS. For intrapsychic foundations, three of the seven items are retained including motivation, anhedonia, and capacity for empathy (items 14, 16, 20). For interpersonal relations, two of eight items are retained, including active acquaintances and social initiative (3, 6). For instrumental role, only one item, occupational role functioning (9), of four is retained. For common objects and activities, one of the two items is retained, possession of objects reflecting participation in living (18). These observable signs provide benchmarks of the degree and extent of the deficit state. They develop over time and enable early detection and therapeutic intervention. Items that do not provide significant additional predictive ability beyond these seven items include intimate relations, social activity, level of accomplishment, degree of underemployment, satisfaction with occupation, sense of purpose, curiosity, time utilization, commonplace activities, and capacity for engagement with the interviewer. Thus, it seems that drive and initiative are more predictive of overall QOL, as measured by the total QLS, than actual accomplishments and satisfaction.

Lack of drive and initiative are part of the deficit state. Deficit psychopathology has been proposed as a separate disease within the schizophrenia syndrome (Kirkpatrick *et al*, 2001) and is associated with poorer QOL, even when considering the severity of negative symptoms at baseline (Tek *et al*, 2001). Our results suggest that focusing on the

assessment of initiative would be a productive target for clinical intervention including rehabilitation efforts.

The purpose of our study was to shorten the QLS, while retaining the predictive power for the total score. The abbreviated scale still incorporates items from all four subscales, which in the original and abbreviated form are not balanced with respect to number of items. Recent publications that have evaluated QOL in schizophrenia, have utilized total QOL scores as outcome measures (Tek *et al*, 2001; Cramer *et al*, 2001, 2000).

There are several limitations to the present study. Most importantly, the initial evaluation is based on the month preceding intake into the Schizophrenia Center. Since the emphasis of this study was the statistical treatment of the scale, we have not examined the relation between symptoms and QOL. Also, we have not evaluated the extent to which QOL relates to neurocognitive deficits or to measures of brain anatomy and function. Furthermore, results may differ when the seven items are administered in isolation and may not provide as much information as the longer scale. However, the brevity of the scale makes it attractive to clinicians who wish to monitor their patients' progress in this crucial domain of care. The importance of improving the QOL of individuals with schizophrenia has gained appreciation among consumers, family members, and health professionals.

REFERENCES

Bow-Thomas CC, Velligan DI, Miller AL, Olsen J (1999). Predicting quality of life from symptomatology in schizophrenia at exacerbation and stabilization. *Psychiatry Res* **86**: 131–142.

Buchanan RW, Holstein C, Breier A (1994). The comparative efficacy and long-term effect of clozapine treatment on neuropsychological test performance. *Biol Psychiatry* **36**: 717–725.

Cramer J, Rosenheck R, Xu W, Henderson W, Thomas J, Charney D (2001). Detecting improvement in quality of life and symptomatology in schizophrenia. *Schizophr Bull* **27**: 227–234.

Cramer JA, Rosenheck R, Xu W, Thomas J, Henderson W, Charney DS (2000). Quality of life in schizophrenia: a comparison of

instruments. Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia. *Schizophr Bull* **26**: 659–666.

Draper NR, Smith H (1998). *Applied Regression Analysis*, 3rd edn. Wiley: New York.

Green MF (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *Am J Psychiatry* **153**: 321–330.

Gur RE, Mozley PD, Resnick SM, Levick S, Erwin R, Saykin AJ *et al* (1991). Relations among clinical scales in schizophrenia: overlap and subtypes. *Am J Psychiatry* **148**: 472–478.

Hamilton SH, Revicki DA, Genduso LA, Beasley (1998). Olanzapine versus placebo and haloperidol: quality of life and efficacy results of the North American double-blind trial. *Neuropsychopharmacology* **18**: 41–49.

Henrichs DW, Hanlon TE, Carpenter WT (1984). The quality of life scale: an instrument for rating the schizophrenic deficit syndrome. *Schizophr Bull* **10**: 388–398.

Hocking RR, Leslie RN (1967). Selection of the best subset in regression analysis. *Technometrics* **9**: 531–540.

Kirkpatrick B, Buchanan RW, Ross DE, Carpenter WT (2001). A separate disease within the syndrome of schizophrenia. *Arch Gen Psychiatry* **58**: 165–171.

Medalia A, Aluma M, Tryon W, Merriam AE (1998). Effectiveness of attention training in schizophrenia. *Schizophr Bull* **24**: 147–152.

Meltzer HY, Burnett S, Bastani B, Ramirez LF (1990). Effects of six months of clozapine treatment on the quality of life of chronic schizophrenic patients. *Hosp Commun Psychiatry* **41**: 892–897.

Rosenheck R, Cramer J, Xu W, Thomas J, Henderson W, Frisman L *et al* (1997). A comparison of clozapine and haloperidol in hospitalized patients with refractory schizophrenia. Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia. *N Engl J Med* **337**: 809–815.

Shrout PE, Fleiss JL (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* **86**: 420–428.

Shtasel DL, Gur RE, Gallacher FG, Heimberg C, Cannon TD, Gur RC (1992). Phenomenology and functioning in first-episode schizophrenia. *Schizophr Bull* **18**: 449–462.

Spitzer RL, Williams TB, Gibbon (1996). *Structured Clinical Interview for the DSM-IV (SCID-P)*. New York State Psychiatric Institute: New York.

Tek C, Kirkpatrick B, Buchanan RW (2001). A five-year followup study of deficit and nondeficit schizophrenia. *Schizophr Res* **49**: 253–260.

Wykes T, Reeder C, Corner J, Williams C, Everitt B (1999). The effects of neurocognitive remediation on executive processing in patients with schizophrenia. *Schizophr Bull* **25**: 291–307.