International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 18(2): 138–146 (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.285

A psychometric evaluation of the clinician-rated Quick Inventory of Depressive Symptomatology (QIDS-C_{16}) in patients with bipolar disorder

IRA H. BERNSTEIN,[1,3] A. JOHN RUSH,[2,3] TRISHA SUPPES,[2] MADHUKAR H. TRIVEDI,[2] ADA WOO,[1] YASUSHI KYUTOKU,[1] M. LYNN CRISMON,[4] ELLEN DENNEHY[5] & THOMAS J. CARMODY[2]

1 The University of Texas, Arlington, TX, USA
2 Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, USA
3 Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA
4 University of Texas College of Pharmacy, Austin, TX, USA
5 Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA

Key words
depression, bipolar disorder, symptom measures, Quick Inventory of Depressive Symptomatology

Correspondence
A. John Rush, Department of Psychiatry, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-9066, USA.
Telephone (+1) 214-648-4832
Fax (+1) 214-648-6878
Email: john.rush@utsouthwestern.edu

Received 13 April 2007;
revised 5 March 2008;
accepted 28 May 2008

Abstract

The clinician-rated, 16-item Quick Inventory of Depressive Symptomatology (QIDS-C_{16}) has been extensively evaluated in patients with major depressive disorder (MDD). This report assesses the psychometric properties of the QIDS-C_{16} in outpatients with bipolar disorder (BD, $N = 405$) and MDD ($N = 547$) and in bipolar patients in the depressed phase only (BD-D) ($N = 99$) enrolled in the Texas Medication Algorithm Project (TMAP) using classical test theory (CTT) and the Samejima graded item response theory (IRT) model.

Values of coefficient alpha were very similar in BD, MDD, and BD-D groups at baseline ($\alpha = 0.80–0.81$) and at exit ($\alpha = 0.82–0.85$). The QIDS-C_{16} was unidimensional for all three groups. MDD and BD-D patients ($n = 99$) had comparable symptom levels. The BD-D patients ($n = 99$) had the most, and bipolar patients in the manic phase had the least depressive symptoms at baseline. IRT analyses indicated that the QIDS-C_{16} was most sensitive to the measurement of depression for both MDD patients and for BD-D patients in the average range.

The QIDS-C_{16} is suitable for use with patients with BD and can be used as an outcome measure in trials enrolling both BD and MDD patients.

Introduction

The Quick Inventory of Depressive Symptomatology (QIDS) includes 16 items that are contained within the 30-item Inventory of Depressive Symptomatology (IDS) (Rush *et al.*, 2000; Rush *et al.*, 2003a; Trivedi *et al.*, 2004a). These items were selected to measure the nine core symptom domains that define a major depressive episode

(DSM-IV) (APA, 2000) including Sleep (four items), Sad mood (one item), Appetite/weight change (four items), Concentration/decision-making (one item), Self view (one item), Thoughts of death or suicide (one item), General interest (one item), Energy level (one item), and Restlessness/agitation (two items) (Rush *et al.*, 1986; Rush *et al.*, 1996). The severity of each of the nine criterion symptom domains is defined by the most pathological response in the item or set of items used to define each domain.

There is considerable evidence for the reliability and validity of the QIDS-C_{16} in patients with major depressive disorder (MDD) (Rush *et al.*, 2003b; Rush *et al.*, 2006a; Trivedi *et al.*, 2004b). The clinician-rated, 16-item Quick Inventory of Depressive Symptomatology (QIDS-C_{16}) can be administered in 5–7 minutes and therefore is cost-efficient. Cost and time efficiency are especially important in the outpatient management of patients with major depressive disorder (MDD) and bipolar disorder (BD) (Chioqueta and Stiles, 2003). The 16-item QIDS comes in three forms: self-report (QIDS-SR_{16}), clinician (QIDS-C_{16}), and interactive voice response (QIDS-IVR_{16}). These three forms yield highly similar results in outpatients with MDD (Rush *et al.*, 2006a). All three versions of the QIDS are also highly sensitive to treatment effects. This report evaluates the QIDS-C_{16} in depressed outpatients with BD.

Several depression rating scales, including the Montgomery–Asberg Depression Rating Scale (MADRS) (Montgomery and Äsberg, 1979; Osterberg and Blaschke, 2005), the Hamilton Rating Scale for Depression (Hamilton, 1960, 1967; Osterberg and Blaschke, 2005), and the Bech–Rafaelsen Melancholia Scale (Bech and Rafaelsen, 1980) have been used to assess depressive symptom severity in patients with BD. The QIDS, however, focuses solely on the nine criterion symptom domains that define a major depressive episode by DSM-IV (APA, 2000). This focus is consistent with the aim of measuring remission (absence of the criterion symptoms for a major depressive episode), as suggested by the recent ACNP Task Force Report on remission (Rush *et al.*, 2006c). Since the QIDS comes in both clinician-rated and self-report versions, it is of interest to know how each of these measures performs in a bipolar sample.

This study addressed two questions: (a) what are the psychometric properties of the QIDS-C_{16} in BD patients and (b) are there differences between BD patients and MDD patients in either the relative frequency of the various symptom domains that are endorsed or in the relation of the individual symptom domains to overall depression?

Methods and materials

Participants, design, and procedure

All subjects in this report had either MDD or Bipolar I Disorder, including schizoaffective disorder-bipolar type. Participants were enrolled in the Texas Medication Algorithm Project (TMAP) (Rush *et al.*, 2003b; Suppes *et al.*, 2003), whether in treatment as usual or the algorithm treatment group, as long as they had both a baseline and at least one post-baseline measure. Patients could be in the depressed, manic, or mixed phase. Patients were diagnosed by a traditional psychiatric clinical interview using DSM-IV criteria (APA, 1980) (see Gilbert *et al.*, 1998; Rush *et al.*, 2003b; Suppes *et al.*, 2003; Trivedi *et al.*, 2004b) conducted by practicing psychiatrists in the public sector. All diagnoses were checked with a checklist used by a Clinical Research Coordinator at each site. No structural interview was used. Subjects were outpatients at least 18 years of age. They had to be sufficiently symptomatic as to require initiation of medication or a change in present type of medication. Most were on medications at baseline. Patients were enrolled at 19 outpatient clinics across Texas, as described in detail elsewhere (Suppes *et al.*, 2003; Trivedi *et al.*, 2004b). Subjects who met criteria for schizoaffective disorder were divided into those with 'bipolar type' (included in the bipolar sample) and those with shizophrenic type (and excluded). This distinction was made based on the history of clear-cut manic or major depressive episodes with or without psychotic symptoms, in which some psychotic symptoms could persist outside the mood episodes (schizoaffective-bipolar type). For the bipolar-schizoaffective type, the affective episodes were mild and indistinct, present on top of a chronically psychotic course of illness. Patients who required inpatient hospitalization for detoxification at study entry, or who were receiving mental retardation service or treatment in an Assertive Community Treatment program, or were not able to give informed consent were excluded. TMAP was conducted in accordance with international guidelines for good clinical practice and the Declaration of Helsinki and approved by the Institutional Review Boards at the University of Texas Southwestern Medical Center at Dallas and the University of Texas at Austin. All subjects provided written, informed consent prior to participation in the study. For these analyses, all enrollees with MDD or BD were included.

The QIDS-C_{16} was obtained by extracting the 16 items that constitute the QIDS-C_{16} from the 30-item clinician-rated Inventory of Depressive Symptomatology (IDS-C_{30}), which was obtained by Research Outcome Assessors (trained counselors, psychologists, social workers, or

nurses) at baseline and every three months thereafter for up to 24 months. The QIDS-C_{16} was scored as previously described (Rush *et al.*, 2003a, 2006b) so that nine symptom domains are each rated 0–3. These nine domains reflect the criterion symptoms required to diagnose a major depressive episode by DSM-IV (APA, 2000). The range is 0–27, with no (0–5), mild (6–10), moderate (11–15), severe (16–20), and very severe (21+) symptom severity.

Statistical methods

The psychometric properties of the QIDS-C_{16} were evaluated at baseline and at exit by classical test theory (CTT) and the Samejima graded item response theory (IRT) methods (Samejima, 1997) as previously described (Rush *et al.*, 2006b). CTT analyses involved calculating the means and standard deviations of the nine symptom domains and computing the domain/total correlations (r_{it}) between these nine domains and the total score (sum of individual domain scores).

IRT analyses involved generating three dichotomies from the four response categories for each item (i.e. 0, 1, 2, 3). For each category, zero denotes the absence of symptomatology and three denotes severe symptomatology. The three dichotomies were (i) 0 versus 1, 2, or 3, (ii) 0 or 1 versus 2 or 3, and (iii) 0, 1, and 2 versus 3. For each domain, the analysis provided a common slope measure (a) and three location measures (b_0, b_1, and b_2). The slope measures, analogous to the item total correlation (r_{it}), describe the strength of relationship between each domain and overall depression. The three location measures, somewhat analogous to item means, describe symptom frequency. In particular, the larger the location estimate, the more often the less pathological categories (e.g. 0, 1 versus 2, 3) are used. For example, higher b_0 values are associated with less frequent endorsement of any of the three pathological categories (1, 2, or 3); the lower the b_0 value, the more commonly the symptom is reported.

IRT analysis entails an explicit method by which to compare diagnostic groups (e.g. BD versus MDD) with respect to one or more domains. This is accomplished by comparing a model in which one or more terms (a, b_0, b_1, and/or b_2) are constrained to equality between the groups with a model in which they are free to vary. A comparison of these two models generates a form of chi-square called G^2. Significance implies that the groups differ with respect to the term(s) studied. Such differences are known as differential item functioning (DIF). Slope differences imply that individual symptom domains relate differently to overall depression for BD and MDD patients. Location differences imply that symptoms are reported with different frequencies between the two groups. These analyses complement the CTT findings of mean differences between groups.

IRT assumes that the nine QIDS domains vary unidimensionally. To establish unidimensionality in this sample, the obtained scree was compared with randomly generated scree values derived from the same number of observations, with a procedure known as parallel analysis (Horn, 1965; Humphreys and Ilgen, 1969; Humphreys and Montanelli, 1975; Montanelli and Humphreys, 1976) for the MDD, BD, and bipolar patients in the depressed phase only (BD-D) groups.

An important part of the IRT analysis is the generation of the test information function (TIF), which describes the sensitivity of the instrument to slight differences in depression over individuals with a range of depression severity as a function of depression level, which serves a similar role as coefficient alpha in CTT, but TIF is a function of depression level rather than an omnibus measure.

IRT analyses were performed using Multilog; the remaining analyses were performed using SAS.

Results

Sample features

Participants included 405 outpatients with BD and 547 outpatients with MDD. In the BD group, patients were in the depressed (BD-D) (99/405 = 24.4%), mixed (111/405 = 27.4%), or manic (113/405 = 27.9%) phases, and the BD group also included subjects with schizoaffective disorder, bipolar type (14.8%; 60/405), and another 5.4% (22/405) in an unspecified phase. Two MDD patients with incomplete QIDS-C_{16} data at exit were deleted from some analyses. The BD and MDD samples were 70% and 79% female, respectively. The BD sample was 9.4% African-American, 27.7% Hispanic, 61.6% White, and 1.2% 'other.' The MDD sample was 20.8% African-American, 26.1% Hispanic, 51.7% White, and 1.3% 'other.' Overall, 19.4% of the MDD sample was psychotic, as was 34.1% of the BD sample. Generally, patients with BD were on public assistance (48.2%) and 81.3% graduated high school or had a General Educational Development (GED) credential. The comparable figures for MDD patients were 45.9% and 66.7% respectively.

Baseline versus exit CTT

Exit data were obtained at 3–24 months after study entry. The mean time from baseline to exit was 13.3 months [standard deviation (SD) = 4.6] for BD and 12.8 months

(SD = 5.0) for MDD. The baseline and exit QIDS-C_{16} means were 12.1 (SD = 5.8) and 10.1 (SD = 5.9), respectively, for the BD and 15.0 (SD = 5.5) and 11.8 (SD = 6.3), respectively for the MDD group. The decline in QIDS-C_{16} scores within each group was significant, $t(404) = 6.88$ (for BD) and $t(544) = 11.27$; $p < 0.01$ (for MDD). As might be expected, the QIDS-C_{16} MDD means were significantly greater than the BD means at both baseline and exit, $t(950) = 7.88$ and $t(949) = 2.94$; $p < 0.01$. The variance also increased within both groups from baseline to exit. The latter increases were associated with slightly increased coefficient alphas from 0.80 to 0.83 among the BD patients and 0.80 to 0.85 among the MDD patients from baseline to exit.

Table 1 CTT findings [item means, item standard deviations (SDs), item/total correlations (r_{it}), and coefficient alpha (α)] for all bipolar, depressed bipolar, and all MDD patients at baseline

	Bipolar[a] (*N* = 405)			BD-D (*N* = 99)			MDD (*N* = 547)		
Domain	Mean	SD	r_{it}	Mean	SD	r_{it}	Mean	SD	r_{it}
1 Sleep	2.26	0.92	0.43	2.39	0.78	0.35	2.45	0.85	0.37
2 Sad mood	1.36	0.99	0.55	1.66	0.93	0.52	1.97	0.94	0.62
3 Appetite	1.75	1.16	0.31	1.97	1.02	0.41	1.97	1.07	0.26
4 Concentration/decision-making	1.33	1.05	0.59	1.75	0.97	0.58	1.61	0.98	0.56
5 Self view	1.12	1.17	0.58	1.57	1.18	0.60	1.46	1.15	0.54
6 Thoughts of death or suicide	0.64	0.85	0.47	0.79	0.86	0.43	0.88	0.92	0.47
7 General interest	1.09	1.11	0.56	1.35	1.08	0.62	1.54	1.11	0.61
8 Energy level	1.22	1.11	0.59	1.59	1.11	0.61	1.67	1.03	0.59
9 Restlessness/agitation	1.32	0.81	0.41	1.44	0.77	0.44	1.44	0.73	0.38
Total	12.1	5.76		14.51	5.55		14.98	5.46	
Coefficient α	0.80			0.81			0.80		

[a] Includes bipolar patients in the depressed (*n* = 99), manic (*n* = 113), mixed (*n* = 111), and unspecified (*n* = 22) phases at baseline as well as 60 schizoaffective patients.

Table 2 CTT findings [item means, item standard deviations (SDs), item/total correlations (r_{it}), and coefficient alpha (α)] for all bipolar, depressed bipolar, and all MDD patients at exit

	Bipolar* (*N* = 405)			BD-D (*N* = 99)			MDD (*N* = 547)		
Domain	Mean	SD	r_{it}	Mean	SD	r_{it}	Mean	SD	r_{it}
1 Sleep	2.29	0.92	0.40	2.52	0.75	0.32	2.27	0.93	0.46
2 Sad mood	1.10	1.00	0.67	1.41	1.01	0.64	1.54	1.04	0.74
3 Appetite/weight	1.40	1.25	0.30	1.86	1.26	0.29	1.46	1.18	0.36
4 Concentration/decision-making	1.07	1.10	0.59	1.35	1.10	0.53	1.27	1.12	0.60
5 Self view	0.85	1.05	0.59	1.16	1.14	0.61	1.06	1.14	0.65
6 Thoughts of death or suicide	0.41	0.72	0.57	0.51	0.84	0.60	0.55	0.77	0.53
7 General interest	0.81	1.12	0.63	0.96	1.15	0.63	1.22	1.17	0.70
8 Energy level	1.01	1.06	0.68	1.29	1.15	0.70	1.31	1.14	0.65
9 Restlessness/agitation	1.15	0.83	0.48	1.24	0.77	0.50	1.16	0.79	0.50
Total	10.10	5.94		12.30	6.00		11.83	6.34	
Coefficient α	0.83			0.82			0.85		

[a] Includes bipolar patients in the depressed (*n* = 99), manic (*n* = 113), mixed (*n* = 111), and unspecified (*n* = 22) phases at baseline as well as 60 schizoaffective patients.

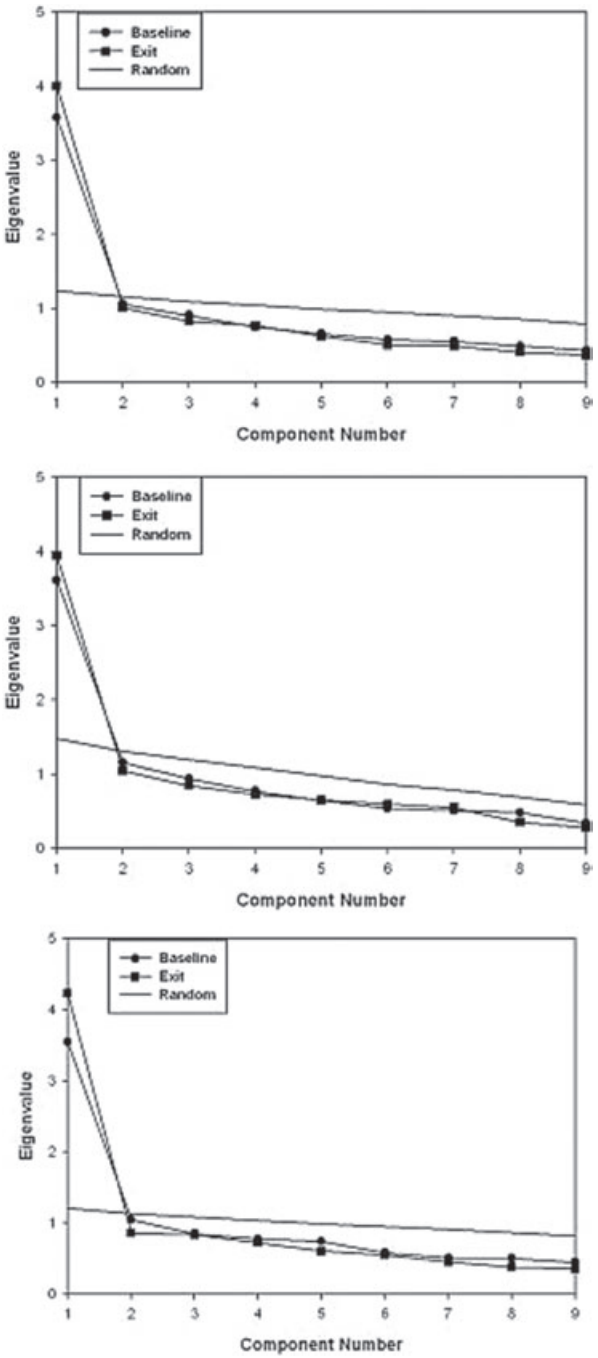

Figure 1 Scree plots of the QIDS-C_{16} for all BD patients (top panel, $N = 405$), BP-D patients (middle panel, $N = 99$), and all MDD patients (lower panel, $N = 547$).

Tables 1 and 2 also show the CTT findings for individual domains. The decline from baseline to exit was found to be significant beyond the 0.01 level for all individual domains except domain 1 (sleep disturbances) with BD patients. At baseline, the MDD patients tended to report greater symptom levels than the BD patients for all domains. At baseline, MDD patients also reported greater symptom levels in all but domains 1, 3, and 9 (Sleep, Appetite, and Restlessness/agitation) as compared to exit.

This analysis was repeated on only the 99 BD-D patients at baseline. Their scale mean was 14.5 (SD = 5.5). Coefficient alpha was 0.81, which was essentially identical to that obtained from the full BD patient sample (0.80). The MDD group averaged 0.5 units higher in their total score at baseline than the BD-D patients, which was non-significant. The only difference between the BD-D and MDD patients in individual QIDS-C_{16} domains was for Sad mood $t(641) = 3.03$, $p < 0.01$. MDD patients tended to report greater levels of sad mood. This analysis could not be performed at exit since exit phase was unknown.

Dimensionality

Figure 1 (top panel) compares the baseline and exit screes obtained from the full BD (top panel) sample with randomly generated scree; the BD-D group (middle panel), and the MDD group (bottom panel). In all groups, the obtained first principal component was much larger than its randomly generated counterpart. The reverse was true of the second component. These results confirm the unidimensionality of the QIDS-C_{16} in the BD, BD-D, and MDD groups at both baseline and exit.

IRT analyses

Slope differences were first assessed by an overall test comparing a model in which all nine pairs of *a* parameters for the BD and MDD groups were allowed to vary freely and a model in which these nine pairs were constrained to equality, allowing all of the 27 intercept parameters (nine domains × three intercepts/domain) to vary freely in both models. The two models did not differ significantly either at baseline or at exit, $G^2(9) = 7.6$ and 8.8. This equality means that each slope can be assumed to have the same value within the two groups, i.e. the domains are equally discriminating within BD and MDD patients. Next, intercept differences were evaluated by comparing a model in which the 27 pairs of intercepts were allowed to vary freely versus a model in which they were constrained to equality, allowing the slopes to vary freely in both cases. Differences were clearly significant at

Int. J. Methods Psychiatr. Res. 18(2): 138–146 (2009). DOI: 10.1002/mpr

baseline and at exit, $G^2(27) = 111.3$ and 90.2; $p < 0.001$, respectively, justifying tests on individual domains.

Differences between BD and MDD intercepts were significant for all individual domains beyond the 0.01 level at baseline save for domains 3 (Concentration/decision making, $p < 0.05$) and 9 [Restlessness/agitation, non-significant (ns)]. At exit, domains 2 (Sad mood), and 8 (Energy level) were significant beyond the 0.01 level. Domains 4 (Concentration/decision making), 5 (Self view), and 6 (Thoughts of death or suicide) were significant beyond the 0.05 level, and domains 1 (Sleep), 3 (Appetite/weight), 7 (General interest), and 9 (Restlessness/agitation) were non-significant. Paralleling the finding that MDD patients tended to report higher levels of symptomatology than BD patients, MDD thresholds were lower in 26 of 27 comparisons at baseline and 24 of comparisons 27 at exit.

When only the BD-D patients were compared to MDD patients, the baseline slope differences were again non-significant, $G^2(9) = 6.1$, but the intercept differences were significant, $G^2(27) = 41.1$, $p < 0.01$. The only baseline intercept to differ significantly was domain 2 (Sad mood, $p < 0.05$). In other words, the rather pronounced threshold differences observed when all BD patients were compared with the MDD patients was due to the fact that most of the BD patients were not in a depressed mood. When the comparison was limited to depressed patients, group differences were minimal.

Test information functions (TIFs)

Figure 2 contains the TIFs for the QIDS-C_{16} at baseline and at exit. Data for the full BD sample (top panel), the depressed phase or BD-D only sample ($n = 99$), and the MDD patients (bottom panel) are shown. The function peaks above the means of BD distributions but near the mean of the MDD patients. This finding reflects the mean difference between BD and MDD groups and suggests that the test is most sensitive to detecting differences for MDD patients at an average level of depression which is similar to BD-D patients.

Subgroup differences within BD sample

Data were available as to phase of illness (depressed, manic, mixed, schizoaffective, or other) for the BD patients, and as to subtype (non-psychotic and psychotic) for the MDD patients. Differences in overall QIDS-C_{16} scores varied significantly among the BD patient groups, $F(4\,404) = 9.45$. The BD-D patients had the highest means;

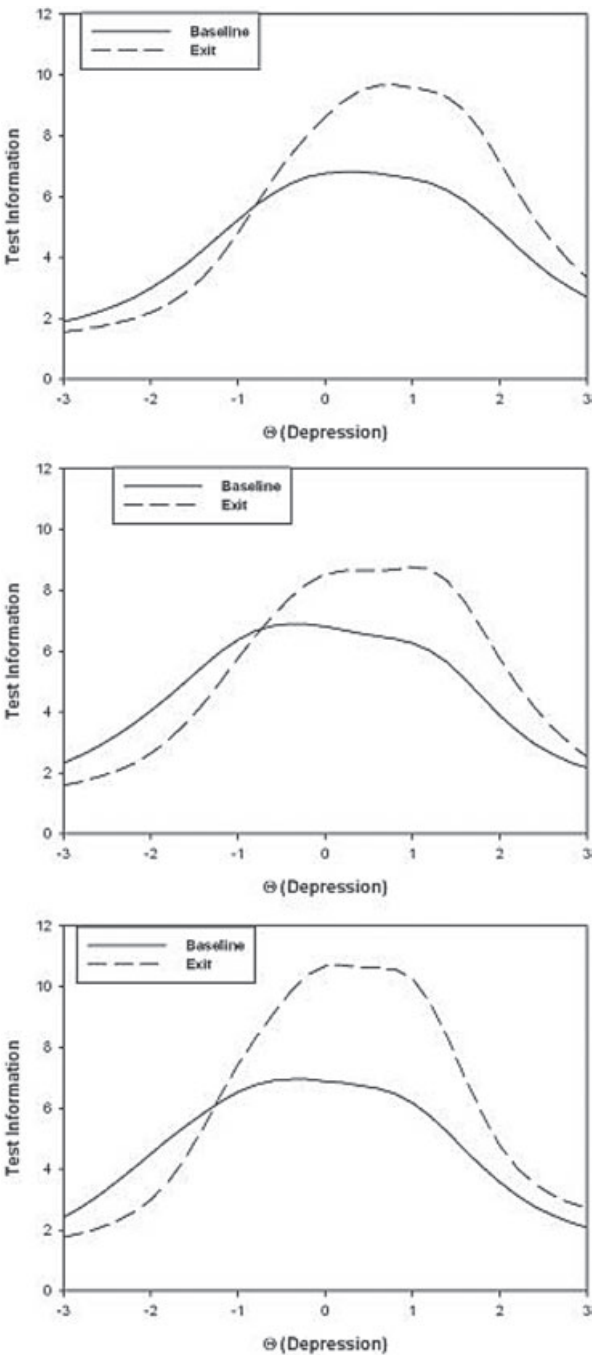

Figure 2 Test information functions for the QIDS-C_{16} at baseline and exit for all BD patients (top panel, $N = 405$), BP-D patients (middle panel, $N = 99$), and all MDD patients (lower panel, $N = 547$).

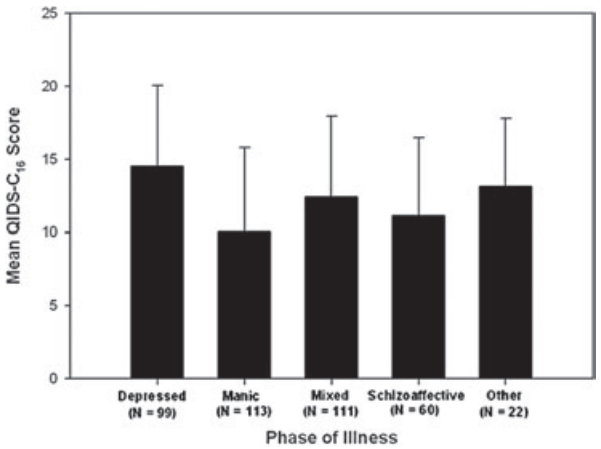

Figure 3 Mean baseline QIDS-C_{16} scores for BD subgroups defined on the basis of phase at baseline; vertical bars denote one standard deviation.

patients in a manic phase had the lowest means. Figure 3 presents these QIDS-C_{16} means +1 standard deviation for each subgroup, respectively. Conversely, psychotic and non-psychotic MDD patients did not differ with respect to their QIDS-C_{16} means, $F(1\,539) = 3.75$.

Discussion

These results indicate that (i) the psychometric properties of the QIDS-C_{16} are satisfactory in BD patients overall and in BD-D (BD-depressed phase only) patients, (ii) the QIDS-C_{16} is unidimensional for BD, BD-D, and MDD groups, and (iii) the relation between these symptom domains to overall depression is highly similar for MDD and BD patients, and were very similar for BD-D and MDD patients. The psychometric properties of the QIDS-C_{16} in the full BD and in the BD-D samples were very similar to those previously found with MDD patients. These results confirm prior findings (Rush *et al.*, 2005; Rush *et al.*, 2006b; Trivedi *et al.*, 2004a) for patients with MDD indicating that the QIDS-C_{16} is a valid measure of depression.

These findings apparently contradict some previous research that has suggested the need for different scales to assess depression severity in each disorder, most specifically Berk *et al.* (2004). However, Berk *et al.*'s (2004) review discussed symptoms other than those constituting the depressive core, e.g. symptoms of atypical depression, and also dealt with other differences between MDD and BD such as age of onset. Our assertion based upon the present data is that one cannot use the core symptoms of depression to differentiate MDD and BD. In the present study, BD patients overall tended to report less symptom severity than MDD patients, but the relation between each domain and overall depression was the same for both groups. That severity is lower in a mixed BD sample is expected. When the BD-D only sample was evaluated, severity was comparable to the MDD sample. Thus, we do not feel that there is need for a scale to measure core depressive symptoms separately for BD patients. Symptoms found in subgroups of patients are a separate issue.

Limitations

One possible limitation is that both TMAP groups are of limited education, so the findings may not generalize to a better educated sample. However, it is just as possible that the limited education may be a benefit since the QIDS-C_{16} may work even better on a better educated BD sample. Our previous findings in patients with MDD showed that the QIDS worked about equally well on a better educated sample (Rush *et al.*, 2005; Rush *et al.*, 2006b; Trivedi *et al.*, 2004a). Whether the same is true for a better educated BD sample remains to be evaluated.

A second limitation is that the QIDS-C_{16} was extracted from the full IDS-C_{30}, so the context provided by those items in the IDS-C_{30} that are not part of the QIDS-C_{16} may have been beneficial to the psychometric properties of the QIDS-C_{16}.

A third limitation is that patients could be taking any medication at baseline or exit, which, in turn, could affect the symptom profiles found. Thus, psychometric assessment in a medication-free bipolar sample has yet to be accomplished.

These results are limited to the QIDS-C_{16}. Whether similar results would pertain to the self-report or interactive voice versions of the QIDS is unknown. However, the results obtained with the QIDS-C_{16} and self-report versions have yielded highly similar results with MDD patients (Rush *et al.*, 2005; Rush *et al.*, 2006a; Rush *et al.*, 2006b; Trivedi *et al.*, 2004b).

Finally, we did not compare the QIDS-C_{16} to other depression scales in this study. Such comparisons are indicated.

Conclusion

These results support the use of the QIDS-C_{16} to assess depression in BD patients. Taken with other reports (Trivedi *et al.*, 2004a), the QIDS-C_{16} can be reliably used in studies of patient samples that include both patients with MDD and BD.

Int. J. Methods Psychiatr. Res. 18(2): 138–146 (2009). DOI: 10.1002/mpr

Acknowledgments

This report was funded in part by the National Institute of Mental Health (NIMH), National Institutes of Health (MH-68851 to the University of Texas Southwestern Medical Center at Dallas, A. John Rush, M.D., PI, and by MH-68852 to the University of Texas at Arlington, Ira H. Bernstein, Ph.D., PI). The original research project was also supported by NIMH Grant MH-53799, the Robert Wood Johnson Foundation, the Meadows Foundation, the Lightner-Sams Foundation, the Nanny Hogan Boyd Charitable Trust, the Texas Department of Mental Health and Mental Retardation, the Center for Mental Health Services, the Betty Jo Hay Distinguished Chair in Mental Health and the Rosewood Corporation Chair in Biomedical Science (AJR), the United States Pharmacopoeia Convention, Inc., Mental Health Connections, a partnership between Dallas County Mental Health and Mental Retardation (MHMR) and the Department of Psychiatry of the University of Texas Southwestern Medical Center, which received funding from the Texas State Legislature and the Dallas County Hospital District and the University of Texas at Austin College of Pharmacy, and the Southwestern Drug Corporation Centennial Fellowship in Pharmacy. The following pharmaceutical companies provided unrestricted educational grants: Abbott Laboratories, AstraZeneca, Bristol-Myers Squibb, Eli Lilly & Company, Forest Laboratories, GlaxoSmithKline, Janssen Pharmaceutica, Novartis, Organon, Pfizer, Inc. and WyethAyerst Laboratories, Inc.

Declaration of interests statement

The authors have no competing interests.

References

American Psychiatric Association (APA). (2000) *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., Text Revision), American Psychiatric Press.

Bech P., Rafaelsen O.J. (1980) The use of rating scales exemplified by a comparison of the Hamilton and the Bech–Rafaelsen Melancholia Scale. *Acta Psychiatrica Scandinavica*, **62**, 128–132.

Berk M., Malhi G.S., Mitchell P.B., Cahill C.M., Carman A. C., Hadzi-Pavlovic D., Hawkins M.T., Tohen M. (2004) Scale matters: The need for a Bipolar Depression Rating Scale (BDRS). *Acta Psychiatrica Scandinavica Supplement*, **422**, 39–45.

Chioqueta A.P., Stiles T.C. (2003) Suicide risk in outpatients with specific mood and anxiety disorders. *Crisis*, **24**, 105–112.

Gilbert D.A., Altshuler K.Z., Rago W.V., Shon S.P., Crismon M.L., Toprac M.G., Rush A.J. (1998) Texas Medication Algorithm Project: Definitions, rationale, and methods to develop medication algorithms. *Journal of Clinical Psychiatry*, **59**, 345–351.

Hamilton M. (1960) A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, **23**, 56–62.

Hamilton M. (1967) Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, **6**, 278–296.

Horn J.L. (1965) An empirical comparison of various methods for estimating common factor scores. *Educational and Psychological Measurement*, **25**, 313–322.

Humphreys L.G., Ilgen D. (1969) Note on a criterion for the number of common factors. *Educational and Psychological Measurement*, **29**, 571–578.

Humphreys L.G., Montanelli R.G. Jr. (1975) An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, **10**, 193–206.

Montanelli R.G. Jr, Humphreys L.G. (1976) Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika*, **41**, 341–348.

Montgomery S.A., Äsberg M. (1979) A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, **134**, 382–389.

Osterberg L., Blaschke T. (2005) Adherence to medication. *New England Journal of Medicine*, **353**, 487–497.

Rush A.J., Carmody T.J., Reimitz P.E. (2000) The Inventory of Depressive Symptomatology (IDS): Clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. *International Journal of Methods in Psychiatric Research*, **9**, 45–59, DOI: 10.1002/mpr.79.

Rush A.J., Gullion C.M., Basco M.R., Jarrett R.B., Trivedi M.H. (1996) The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine*, **26**, 477–486.

Rush A.J., Giles D.E., Schlesser M.A., Fulton C.L., Weissenburger J., Burns C. (1986) The Inventory for Depressive Symptomatology (IDS): Preliminary findings. *Psychiatry Research*, **18**, 65–87.

Rush A.J., Carmody T.J., Ibrahim H.M., Trivedi M.H., Biggs M.M., Shores-Wilson K., Crismon M.L., Toprac M.G., Kashner T.M. (2006a) Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. *Psychiatric Services*, **57**, 829–837. DOI: 10.1176/appi.ps.57.6.829.

Rush A.J., Bernstein I.H., Trivedi M.H., Carmody T.J., Wisniewski S., Mundt J.C., Shores-Wilson K., Biggs M.M., Woo A., Nierenberg A.A., Fava M. (2006b) An evaluation of the Quick Inventory of Depressive Symptomatology and the Hamilton Rating Scale for Depression: A sequenced treatment alternatives to relieve depression trial report. *Biological Psychiatry*, **59**, 493–501, DOI: 10.1016/j.biopsych.2005.08.022.

Rush A.J., Trivedi M.H., Ibrahim H.M., Carmody T.J., Arnow B., Klein D.N., Markowitz J.C., Ninan P.T., Kornstein S., Manber R., Thase M.E., Kocsis J.H., Keller M.B. (2003a) The 16-item Quick Inventory of Depressive

Symptomatology (QIDS), clinician-rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, **54**, 573–583, Erratum p. 585, DOI: 10.1016/S0006-3223(02)01866-8.

Rush A.J., Crismon M.L., Kashner T.M., Toprac M.G., Carmody T.J., Trivedi M.H., Suppes T., Miller A.L., Biggs M.M., Shores-Wilson K., Witte B.P., Shon S.P., Rago W. V., Altshuler K.Z. (2003b) Texas Medication Algorithm Project, phase 3 (TMAP-3): Rationale and study design. *Journal of Clinical Psychiatry*, **64**, 357–369.

Rush A.J., Kraemer H.C., Sackeim H.A., Fava M., Trivedi M.H., Frank E., Ninan P.T., Thase M.E., Gelenberg A.J., Kupfer D.J., Regier D.A., Rosenbaum J.F., Ray O., Schatzberg A.F. (2006c) Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology*, **31**, 1841–1853.

Rush A.J., Trivedi M.H., Carmody T.J., Ibrahim H.M., Markowitz J.C., Keitner G.I., Kornstein S.G., Arnow B., Klein D.N., Manber R., Dunner D.L., Gelenberg A.J., Kocsis J.H., Nemeroff C.B., Fawcett J., Thase M.E., Russell J.M., Jody D.N., Borian F.E., Keller M.B. (2005) Self-reported depressive symptom measures: Sensitivity to detecting change in a randomized, controlled trial of chronically depressed, nonpsychotic outpatients. *Neuropsychopharmacology*, **30**, 405–416, DOI: 10.1038/sj.npp.1300614.

Samejima F. (1997) Graded response model. In: *Handbook of Modern Item Response Theory* (eds van Linden W., Hambleton R.K.), pp. 85–100, Springer-Verlag.

Suppes T., Rush A.J., Dennehy E.B., Crismon M.L., Kashner T.M., Toprac M.G., Carmody T.J., Brown E.S., Biggs M. M., Shores-Wilson K., Witte B.P., Trivedi M.H., Miller A.L., Altshuler K.Z., Shon S.P. (2003) Texas Medication Algorithm Project, phase 3 (TMAP-3): Clinical results for patients with a history of mania. *Journal of Clinical Psychiatry*, **64**, 370–382.

Trivedi M.H., Rush A.J., Ibrahim H.M., Carmody T.J., Biggs M.M., Suppes T., Crismon M.L., Shores-Wilson K., Toprac M.G., Dennehy E.B., Witte B., Kashner T.M. (2004a) The Inventory of Depressive Symptomatology, clinician rating (IDS-C) and self-report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, clinician rating (QIDS-C) and self-report (QIDS-SR) in public sector patients with mood disorders: A psychometric evaluation. *Psychological Medicine*, **34**, 73–82.

Trivedi M.H., Rush A.J., Crismon M.L., Kashner T.M., Toprac M.G., Carmody T.J., Key T., Biggs M.M., Shores-Wilson K., Witte B., Suppes T., Miller A.L., Altshuler K.Z., Shon S.P. (2004b) Clinical results for patients with major depressive disorder in the Texas Medication Algorithm Project. *Archives of General Psychiatry*, **61**, 669–680.