Available online at www.sciencedirect.com
ScienceDirect
Behavior Therapy 41 (2010) 423–431
Behavior Therapy
www.elsevier.com/locate/bt

# Using the QIDS-SR$_{16}$ to Identify Major Depressive Disorder in Primary Care Medical Patients

Brittain E. Lamoureux
Eftihia Linardatos
David M. Fresco
Kent State University

Dena Bartko
Everett Logue
Lori Milo
Summa Health System, Department of Family Medicine

**Major depressive disorder (MDD) is a serious and prevalent mental health issue. As the majority of MDD cases are identified and treated by one's primary care physician, it is imperative that care providers utilize accurate and efficient methods for diagnosing MDD in primary care settings. The present study is the first to investigate the accuracy of the Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR$_{16}$) as a screen for MDD. A heterogeneous sample of 155 primary care patients completed the QIDS-SR$_{16}$ prior to attending a primary care appointment. Participants were then assessed for psychopathology using the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID) by clinicians who were blind to QIDS-SR$_{16}$ scores. Scores on the QIDS-SR$_{16}$ were compared to clinician-assessed current and lifetime diagnoses derived from the SCID, which represented the gold-standard criterion measure. Receiver operator characteristic analysis was utilized to determine the optimal QIDS-SR$_{16}$ cut score to correctly classify participants based on their MDD status as assessed by the SCID. The test revealed a robust area under the curve (.82, $p<0.00001$) and suggested that a cut score of 13 or 14 provided the best balance of sensitivity (76.5%) and specificity (81.8%) in this primary care sample. Over 80% of participants were correctly classified. Separate analyses by race were conducted to address the possibility that different cut scores may be more accurate for African American and Caucasians. Findings from the present study provide support for the use of the QIDS-SR$_{16}$ as a screening measure for identifying primary care patients who will meet diagnostic criteria for MDD based on clinician assessment.**

Support for this research was provided to David M. Fresco by the Ohio Board of Regents.

Address correspondence to David M. Fresco, Ph.D., Department of Psychology, Kent State University, 226 Kent Hall Annex, Kent, OH, 44242, USA; e-mail: fresco@kent.edu.

0005-7894/10/423–431/$1.00/0



Major depressive disorder (MDD) is a serious, prevalent, and potentially debilitating mental health issue. Approximately 13% of men and 21% of women in Western, industrialized nations will suffer from MDD in their lifetimes (Kessler et al., 2005a), with 1-year prevalence rates ranging from 6.6% to 10.3% (Kessler, Chiu, Demler, Merikangas, & Walters, 2005b). Due to a variety of factors, including the stigma associated with mental health treatment, the majority of MDD cases in the United States are identified and treated by one's primary care physician (Regier et al., 1993). Screening for MDD in primary care can substantially improve patient outcomes, particularly when combined with efforts to promote adequate treatment and follow-up (Pignone et al., 2002). Thus, accurate and efficient screening measures are needed to alert service providers to further evaluate patients who are likely to have MDD.

The Quick Inventory of Depressive Symptomatology–Self Report (QIDS-$SR_{16}$; Rush et al., 2003) is a brief self-report measure designed to assess depressive symptom severity. The QIDS-$SR_{16}$ requires minimal training to administer and is freely available for use (see www.ids-qids.org), and thus it is both time and resource efficient. The QIDS-$SR_{16}$ has performed well alongside well-established measures of depression in large-scale Sequenced Treatment Alternatives to Relieve Depression (STAR*D; Gaynes et al., 2008; Rush et al., 2006; Rush et al., 2008) clinical trials. Although the QIDS-$SR_{16}$ has been evaluated as an outcome measure to assess changes in the severity of recent depressive symptoms due to treatment, it may also be useful as a screen for MDD and has not yet been evaluated in this capacity. The present study sought to determine the ability of the QIDS-$SR_{16}$ to accurately differentiate primary care medical patients with current MDD from a heterogeneous sample of primary care patients who vary as to their history of current and lifetime clinician-assessed psychopathology. Scores on the QIDS-$SR_{16}$ were compared to clinician-assessed current and lifetime diagnoses derived from the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID; First, Spitzer, Gibbon, & Williams, 2002), which represented the gold-standard criterion measure. Also, other research has indicated that care providers are less likely to detect mental illness in African Americans, perhaps due to a discrepancy between symptoms of psychopathology endorsed in self-report measures versus clinical interview (e.g., Borowsky et al., 2000; Fresco et al., 2001). To address this issue, separate analyses were conducted to investigate whether use of a different cut score was more accurate in detecting MDD for African Americans than for Caucasians.

## Method

### PARTICIPANTS

Participants were 155 patients (123 females) at an urban, public hospital family medical center serving a lower-income population in a midsized midwestern U.S. city. Participants were considered eligible if they were over 18 years of age at the time of interview. The mean age of participants was 39 years ($SD = 14.18$), ranging from 18 to 79 years of age. A similar number of participants self-identified as either Caucasian (74; 47.7%) or African American (68; 43.9%), with the remaining (13; 8.39%) participants identifying as another ethnicity. The mean education level for participants was high school graduate (12.08 years; $SD = 1.95$), with a range of 7 to 18 years of formal education completed. Forty-six participants (30%) were working outside the home at the time of the study.

### MEASURES

*Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR16; Rush et al., 2003)*

The QIDS-$SR_{16}$ (see www.ids-qids.org) is a brief, self-report version of the Inventory of Depressive Symptomatology (IDS; Rush et al., 1986; Rush, Gullion, Basco, Jarrett, & Trivedi, 1996). As with the IDS, both clinician-administered (QIDS-$C_{16}$) and self-report forms (QIDS-$SR_{16}$) are available. The measure consists of 16 items that assess the nine symptom domains used to diagnose a major depressive episode in the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*; American Psychiatric Association, 1994): sleep disturbance (items 1–4), depressed mood (item 5), change in appetite or weight (items 6–9), concentration (item 10), self-criticism (item 11), suicidal ideation (item 12), interest (item 13), energy/fatigue (item 14), and psychomotor agitation/retardation (items 15 and 16). It lacks items included in the IDS concerning atypical and melancholic specifiers and associated symptoms. The responses for each item range from 0 to 3, with 0 indicating the absence of that symptom in the past week. The scoring scheme involves adding the scores for the nine symptom domains to yield a total score, ranging from 0 to 27 (Rush, Carmody, & Reimitz, 2000; available at www.ids-qids.org/Scoring_Instructions.pdf). For the symptom domains that are addressed in more than one item, the domain item with the highest score is used in total score calculation. Depression severity can be estimated from QIDS-$SR_{16}$ scores using the following guidelines: none (0–5), mild (6–10), moderate (11–15), severe (16–20), and very severe ($\geq 21$) (Rush et al., 2003).

The QIDS-$SR_{16}$ has exhibited high internal consistency, ranging from .73 to .92 (Rush et al., 2003), and was .86 for this sample. It has primarily been utilized to detect treatment change and has demonstrated high sensitivity in this capacity (Corruble, Legrand, Zvenigorowski, Duret, & Guelfi, 1999; Gullion & Rush, 1998; Rush, Gullion, Basco, Jarrett, & Trivedi, 1996; Rush et al., 2000, 2003, 2006; Trivedi et al., 2004). As such, this measure has not been evaluated for retest reliability and would not be expected to demonstrate high test-retest reliability insofar as symptoms of depression vary and the instruction set asks respondents to focus on the past week. Likewise, no data are available which may speak to the ability of the QIDS-$SR_{16}$ to discriminate symptoms of depression from other, unrelated constructs. Concurrent

and convergent validity of the IDS and QIDS have been established with well-known measures of depression and total score correlations between scales ranged from .81 to .95 and score conversions have been developed among these measures (Corruble et al., 1999; Rush et al., 1996, 2003).

*Structured Clinical Interview for DSM-IV-TR (SCID; First et al., 2002)*

The SCID is a clinician-administered diagnostic interview for Axis I disorders based on DSM-IV-TR criteria. The SCID is sensitive to both current and past mental health functioning and is well equipped to address issues of comorbidity and differential diagnosis. Primacy of diagnoses is typically assigned based on degree of functional impairment for each disorder. The length of the total interview varied by individual, as a short screen of yes or no questions was used to determine for which modules a particular individual should be assessed. Questions related to current (i.e., within the past month) and past (i.e., greater than 1 month ago) mood disorders were asked of all individuals. The instructions for both the current and past MDD sections required that interviewers establish specific time frames to which participant responses referred. If there were multiple periods during which the participant had experienced depressive symptoms in the past, the same questions were asked relative to each specific period of time. Together, the sections related to current and past symptoms of MDD allowed for a continuous and complete assessment of an individual's experiences of depression throughout his or her lifetime.

Interrater reliability for diagnosis of MDD using the SCID typically ranges from .61 to .80 (Zanarini et al., 2000). In the present study, interrater reliability was established by having three interviewers rate each of 10 randomly selected SCID interviews. The original interviewer would audio record and rate the SCID, then two other study interviewers who were blind to the original ratings would independently rate the SCID. Cohen's kappa ($\kappa$) was calculated by comparing each of the two additional ratings to the original rating, yielding two $\kappa$ values for each interview. For this sample, interrater reliability across interviewers for the SCID ranged from 0.79 to 0.87 (average $\kappa = 0.83$) for all diagnoses and from 0.83 to 0.85 (average $\kappa = 0.84$) for MDD in particular. Any discrepancies in diagnoses were addressed in case conference to establish consensus.

## PROCEDURE

Participants were recruited from the waiting room of the family medical center prior to attending primary care appointments. Each participant was individually approached and screened for eligibility by a research assistant, and the majority of those approached elected to participate (approximately 85%). Participants provided informed consent and then completed a self-report packet containing demographic questions, the QIDS-$SR_{16}$, and other items not relevant to this study's aims. Research assistants were available to answer any questions that arose.

Of the 362 participants who completed the QIDS-$SR_{16}$, 155 (42.8%) elected to complete the SCID interview. The only significant difference between participants who completed the SCID and participants who did not was that participants who were either engaged as homemakers or employed outside the home were less likely to complete the SCID than participants who were retired, disabled, or unemployed. The SCID was administered by one of 6 clinical psychology graduate students who were kept blind to participants' QIDS-$SR_{16}$ scores. Interviewers received extensive training in conducting the interview, including practicum training on administration techniques and 1 year of supervision rounds to establish accuracy of diagnoses. The average number of interviews conducted per interviewer was 26. Participants' physicians were notified if their responses to either the self-report items or SCID interview indicated the presence of MDD. Participants were paid $25 for their time and effort.

Participants chose to complete the SCID interview at that time (16; 10.3%), at a future appointment (117; 75.5%), or at a future time over the telephone (22; 14.2%). Although the SCID was not designed for telephone administration, it has been shown to be acceptably consistent ($\kappa = .66$) with in-person administration for determining a diagnosis of MDD (Cacciola, Alterman, Rutherford, McKay, & May, 1999). The average time between participant completion of the QIDS-$SR_{16}$ and the SCID interview was 27 days ($SD = 39.46$). The majority of participants (107; 69.0%) were able to complete the SCID interview within 3 weeks of completing the QIDS-$SR_{16}$. For these participants, data from the current MDD section of the SCID were used to determine if each met criteria for MDD at the time of QIDS-$SR_{16}$ completion. The remaining participants (48; 31.0%) completed the SCID interview more than 3 weeks after completing the QIDS-$SR_{16}$. For these participants, data from the past MDD section of the SCID were used to determine if each met criteria for MDD at the time of QIDS-$SR_{16}$ completion. If participants were unable to provide a specific time frame for depressive symptoms with relative confidence,

they were not included in study analyses. We also investigated whether the amount of time between QIDS-$SR_{16}$ and SCID completion affected the likelihood of agreement between the two measures as to MDD status. The correlation was not statistically significant ($r = .08$; $p = .35$), suggesting that the time delay between the two administrations did not introduce significant error into our results.

### ANALYSIS PLAN

The current investigation utilized receiver operator characteristic (ROC; Kraemer, 1992) analysis to determine the optimal QIDS-$SR_{16}$ cut score to correctly classify participants based on their MDD status as assessed by the SCID. Analyses for the current study were conducted using STATA 7 (StataCorp, 2001). Specifically, we regressed the dichotomous MDD status variable onto the continuous QIDS-$SR_{16}$ score using a logistic regression procedure. We then saved the predicted probabilities (a log odds transformation of the QIDS-$SR_{16}$ scores) from this analysis as a new variable. After solving the regression equation, the *lroc* command was implemented to derive the area under the curve (AUC) and ROC curve. To test whether the AUC was significantly better than chance, we used the *brier* procedure to generate several decompositions of the Brier mean probability score based on a binary outcome (MDD Status) and a forecast variable (predicted probability of QIDS-$SR_{16}$ scores). Finally, optimal cut scores were obtained using the *roctab* procedure, which generates sensitivity, specificity, and percent correctly classified for each value of the forecast variable. We elected to present three cut scores: optimal sensitivity (the score that optimized sensitivity without reducing specificity to less than chance); optimal specificity (the score that optimized specificity without reducing sensitivity to less than chance); and optimal sensitivity and specificity (the score that produced the best balance of sensitivity and specificity). For each of the reported cut scores, we also provide the positive predictive value (PPV; the proportion of participants with positive test results who were correctly diagnosed) and the negative predictive value (NPV; the proportion of participants with negative test results who were correctly diagnosed).

## Results

### DIAGNOSTIC PREVALENCES

The mean QIDS-$SR_{16}$ score for this sample was 10 ($SD = 6$), indicating a mild to moderate level of depressive symptoms. Based on the recommended thresholds to estimate depression severity (Rush et al., 2003), 41 participants (26.5%) endorsed no depression, 50 participants (32.3%) endorsed mild depression, 32 participants (20.6%) endorsed moderate depression, 24 participants (15.5%) endorsed severe depression, and 8 participants (5.2%) endorsed very severe depression. Based on the SCID, 34 (21.9%) participants met criteria for MDD, 58 (37.4%) participants met criteria for a *DSM-IV-TR* diagnosis other than MDD, and the remaining 63 (40.7%) participants did not meet criteria for a current psychiatric diagnosis, though

Table 1
Diagnostic Prevalences for Current DSM-IV-TR Axis I Disorders Based on the SCID (N = 155)

| DSM-IV Diagnosis (Dx) | Primary Dx | Secondary Dx | Tertiary+ Dx | Total |
|---|---|---|---|---|
| Major Depressive D/O | 29 | 5 | 0 | 34 (21.9%) |
| Bipolar I D/O | 6 | 0 | 0 | 6 (3.9%) |
| Dysthymic D/O | 15 | 1 | 2 | 18 (11.6%) |
| Alcohol Abuse | 1 | 2 | 1 | 4 (2.6%) |
| Alcohol Dependence | 1 | 1 | 2 | 4 (2.6%) |
| Substance Abuse | 2 | 2 | 2 | 6 (3.9%) |
| Substance Dependence | 4 | 3 | 4 | 11 (7.1%) |
| Panic D/O | 4 | 4 | 1 | 9 (5.8%) |
| Specific Phobia | 3 | 4 | 9 | 16 (10.3%) |
| Social Phobia | 1 | 3 | 3 | 7 (4.5%) |
| Posttraumatic Stress D/O | 7 | 4 | 2 | 13 (8.4%) |
| Obsessive-Compulsive D/O | 0 | 0 | 3 | 3 (1.9%) |
| Generalized Anxiety D/O | 10 | 19 | 6 | 35 (22.6%) |
| Anxiety NOS | 1 | 2 | 0 | 3 (1.9%) |
| Binge Eating D/O | 0 | 1 | 2 | 3 (1.9%) |
| Any Axis I Diagnosis | 84 (54.2%) | 51 (32.9%) | 37 (23.9%) | |
| No Diagnosis | – | – | – | 63 (40.7%) |

*Note.* D/O = Disorder, NOS = Not Otherwise Specified. Only disorders for which more than two participants met criteria are included. Many participants had multiple SCID diagnoses, thus the total percentage is greater than 100%.

they may have reported previous psychopathology (see Table 1). See Table 2 for information on comorbid diagnoses for the 34 participants with SCID diagnoses of MDD. Univariate analyses (chi-square and *t* statistics) revealed no significant differences in race, age, or years of education based on SCID MDD diagnosis. Women were somewhat more likely to be diagnosed with MDD than men, $\chi^2(1, N=155)=3.72$, $p=.05$, yet this finding should be interpreted cautiously given the relatively small number of male participants ($n=32$).

### ACCURACY OF THE QIDS-SR$_{16}$ IN SCREENING FOR DEPRESSION

The results of the ROC analysis revealed a robust area under the curve (AUC=.82, $p<0.00001$), indicating that participant QIDS-SR$_{16}$ scores predict a diagnosis of MDD significantly better than chance (see Fig. 1). Findings revealed that a cut score of 8 maximized sensitivity (85.3%) while maintaining specificity above chance (55.4%) and resulted in the correct classification for 29 of the 34 participants with MDD (85.3%) and 67 of the 121 participants without MDD (55.4%). Overall, 61.9% of the sample was correctly classified (PPV=34.9%; NPV=93.1%). A cut score of 16 maximized specificity (88.4%) while maintaining sensitivity above chance (52.9%) and resulted in the correct classification for 18 of the 34 participants with MDD (52.9%) and 107 of the 121 participants without MDD (88.4%). Overall, 80.7% of the sample was correctly classified (PPV=56.3%; NPV=87.0%). Analyses indicated that results for a cut score of 13 or 14 were indistinguishable. Use of either score provided the best overall balance of sensitivity (76.5%) and specificity (81.8%) in this primary care sample, correctly classifying 26 of the 34 MDD participants (76.5%) and 99 of the 121 non-MDD participants (81.8%). Overall, 80.7% of participants were correctly classified as to MDD diagnosis (PPV=54.2%; NPV=92.5%). Findings for all scores obtained on the QIDS-SR$_{16}$ are presented in Table 3.

Table 2
Rates of Comorbidity by Disorder for Participants with SCID diagnoses of MDD (N=34)

| Comorbid Diagnosis | N (%) |
|---|---|
| Generalized Anxiety Disorder | 17 (50.0%) |
| Dysthymic Disorder | 3 (8.8%) |
| Posttraumatic Stress Disorder | 3 (8.8%) |
| Substance Dependence | 3 (8.8%) |
| Specific Phobia | 3 (8.8%) |
| Panic Disorder | 2 (5.9%) |
| Social Phobia | 2 (5.9%) |
| Alcohol Abuse | 1 (2.94%) |
| Alcohol Dependence | 1 (2.94%) |
| Obsessive-Compulsive Disorder | 1 (2.94%) |
| Substance Abuse | 1 (2.94%) |
| No Comorbid Diagnosis | 11 (32.4%) |

*Note.* Only comorbid disorders for which at least one participant met criteria are included. Many participants had multiple comorbid diagnoses, thus the total percentage is greater than 100%.

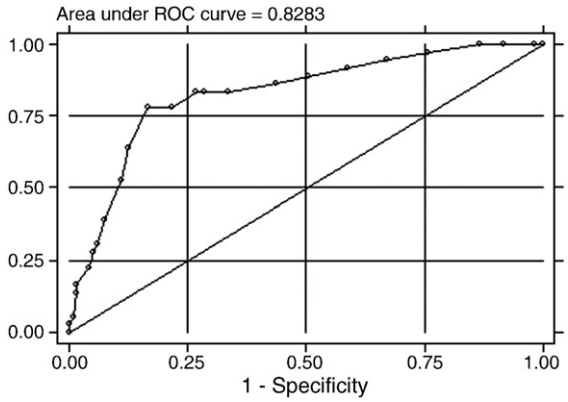


FIGURE 1 Receiver Operator Characteristics (ROC) curve showing the probability of predicting a SCID diagnosis of Major Depressive Disorder using QIDS-SR$_{16}$ score.

Eight (23.5%) of the 34 participants who received SCID diagnoses of MDD scored below a cut score of 13 or 14 on the QIDS-SR$_{16}$. These "false negatives" indicate the degree of under-diagnosis for this sample. Two (25%) of these "false negatives" had SCID diagnoses of MDD judged to be in partial remission. Alternatively, 22 (18.2%) of the 121 participants who did not receive SCID diagnoses of MDD scored at or above a cut score of 13 or 14 on the QIDS-SR$_{16}$. These "false positives" indicate the degree of overdiagnosis for this sample. Eight (36.4%) of these participants met SCID criteria for another mood disorder: 6 (27.3%) Dysthymic Disorder, 1 (4.55%) Bipolar I Disorder, and 1 (4.6%) Depression Not Otherwise Specified (NOS). Six participants (27.3%) did not meet criteria for a current mood disorder, but did meet SCID criteria for another *DSM-IV-TR* Axis I diagnosis. Of the 8 (36.4%) participants who did not meet criteria for any current *DSM-IV-TR* Axis I diagnosis, 4 (18.2%) participants met SCID criteria for past MDD.

The area under the ROC curve was also calculated for this sample excluding participants with no diagnosis ($n=91$). Even using this more conservative estimate, the QIDS-SR$_{16}$ remained a strong predictor of MDD diagnosis (AUC=.75, $p<0.00001$). For this subsample, a cut score of 13, versus 14, performed slightly better at maximizing both sensitivity (76.5%, versus 61.8%) and specificity (66.7%, versus 77.2%). Using 13 as a cut score, 26 of the 34 MDD participants (76.5%) and

Table 3
QIDS-$SR_{16}$ Results by Score (N = 155)

| Score | Sn | Sp | % correctly classified | PPV | NPV | FPR | FNR |
|---|---|---|---|---|---|---|---|
| 1 | 100% | 0% | 21.9% | 21.9% | – | 100% | 0% |
| 2 | 100% | 1.7% | 23.2% | 22.2% | 100% | 98.4% | 0% |
| 3 | 100% | 8.3% | 28.4% | 23.5% | 100% | 91.7% | 0% |
| 4 | 100% | 13.2% | 32.3% | 24.5% | 100% | 86.8% | 0% |
| 5 | 94.1% | 32.2% | 45.8% | 28.1% | 95.1% | 67.8% | 4.9% |
| 6 | 91.2% | 40.5% | 51.6% | 30.1% | 94.2% | 59.5% | 8.8% |
| 7 | 88.2% | 48.8% | 57.4% | 32.6% | 93.7% | 51.2% | 11.8% |
| 8 | 85.3% | 55.4% | 61.9% | 34.9% | 93.1% | 44.6% | 14.7% |
| 9 | 82.4% | 65.3% | 69.0% | 40.0% | 92.9% | 34.7% | 17.7% |
| 10 | 82.4% | 70.3% | 72.9% | 43.8% | 93.4% | 29.8% | 17.7% |
| 11 | 82.4% | 70.3% | 72.9% | 43.8% | 93.4% | 29.8% | 17.7% |
| 12 | 76.5% | 76.9% | 76.8% | 48.2% | 92.1% | 23.1% | 23.5% |
| 13 | 76.5% | 81.8% | 80.7% | 54.2% | 92.5% | 18.2% | 23.6% |
| 14 | 76.5% | 81.8% | 80.7% | 54.2% | 92.5% | 18.2% | 23.6% |
| 15 | 61.8% | 86.0% | 80.7% | 55.3% | 88.9% | 14.1% | 38.2% |
| 16 | 52.9% | 88.4% | 80.7% | 56.3% | 87.0% | 11.6% | 47.1% |
| 17 | 29.4% | 93.4% | 79.4% | 55.6% | 82.5% | 6.6% | 70.6% |
| 18 | 29.4% | 93.4% | 79.4% | 55.6% | 82.5% | 6.6% | 70.6% |
| 19 | 20.6% | 95.0% | 78.7% | 53.9% | 81.0% | 5.0% | 79.4% |
| 20 | 14.7% | 97.5% | 79.4% | 62.5% | 80.3% | 2.5% | 85.3% |
| 21 | 14.7% | 98.4% | 80.0% | 71.4% | 80.4% | 1.7% | 85.3% |
| 23 | 14.7% | 98.4% | 80.0% | 71.4% | 80.4% | 1.7% | 85.3% |
| 25 | 2.9% | 100% | 78.7% | 100% | 78.6% | 0% | 97.1% |
| 26 | 0% | 100% | 78.1% | – | 78.1% | 0% | 100% |

*Note.* Sn = sensitivity, Sp = specificity, % correctly classified, PPV = positive predictive power, NPV = negative predictive power, FPR = false positive rate, FNR = false negative rate. Only the total scores which were represented by at least one participant are included.

38 of the 57 non-MDD participants (66.7%) were correctly classified. Using 14 as a cut score, 21 of the 34 MDD participants (61.8%) and 44 of the 57 non-MDD participants (77.2%) were correctly classified. Overall, 70.3% of participants were correctly classified using 13 as the cut score for MDD diagnosis (PPV = 57.8%; NPV = 82.6%), versus 71.4% using 14 as the cut score (PPV = 61.8%; NPV = 77.2%).

We ran additional analyses to determine the ability of the QIDS-$SR_{16}$ to accurately differentiate participants as to MDD status across race. For Caucasian participants ($n = 74$), the QIDS-$SR_{16}$ significantly predicted MDD diagnosis (AUC = .79, $p = 0.001$). However, a cut score of 13, versus 14, performed slightly better maximizing both sensitivity (78.6%, versus 57.1%) and specificity (81.7%, versus 83.3%). Overall, 81.08% of Caucasian participants were correctly classified using a cut score of 13 (PPV = 50.0%; NPV = 94.2%), versus 78.4% using 14 as the cut score (PPV = 44.4%; NPV = 89.3%). For African American participants ($n = 68$), the QIDS-$SR_{16}$ significantly predicted MDD diagnosis (AUC = .80, $p < 0.01$) and either 10 or 11 as the cut score maximized both sensitivity (78.6%) and specificity (74.1%). Overall, 75.0% of African American participants were correctly classified (PPV = 44.0%; NPV = 93.0%). We further examined the consequences of using a cut score of 13 or 14 for African American participants in this sample, as those were the cut scores of choice based on other analyses. Using a cut score of 13 resulted in a greater number of African American participants correctly classified (79.4%; PPV = 50.0%; NPV = 90.0%) and improved specificity (83.3%), though at the cost of decreased sensitivity (64.3%). The same was true, and more so, for using a cut score of 14. As compared to using 10 or 11 as a cut score, both the percent of African American participants correctly classified (83.8%; PPV = 61.5%; NPV = 89.1%) and specificity (90.7%) increased and sensitivity (57.1%) decreased. We did not have enough participants of other ethnic groups to examine them separately.

## Discussion

Findings from the present study provide initial support for the use of the QIDS-$SR_{16}$ as a screening measure for identifying individuals who meet diagnostic criteria for MDD based on clinician assessment. Using a cut score of 13 or 14 on the QIDS-$SR_{16}$, either of which optimized both the

sensitivity and specificity of the measure, over 80% of participants were correctly classified with regards to MDD status. This level of prediction is significantly better than chance. Importantly, 22 (18.2%) of the 121 participants not diagnosed with MDD using either 13 or 14 as a QIDS-$SR_{16}$ cut score warranted this diagnosis according to the SCID. Likewise, 8 (23.5%) of the 34 participants who were diagnosed with MDD using these QIDS-$SR_{16}$ cut scores should not have received this diagnosis according to the SCID. Although these misdiagnosed participants are clearly in the minority, these errors could have unfortunate consequences for individuals who either go without appropriate mental health treatment or receive unnecessary treatment. These discrepancies point to the need for follow-up evaluation beyond the QIDS-$SR_{16}$, or any screening tool. There may also be some concern regarding the low PPV (54.2%) found for the QIDS-$SR_{16}$ using these cut scores. Given the high cost of undetected depression in terms of the chronicity of the disorder and increased risk for comorbid physical and mental health problems that accompany depression, we may choose to favor screening measures that initially overdetect the disorder and bear the cost of further assessment (Barkow et al., 2002, 2003; Katon & Ciechanowski, 2002; Katon & Schulberg, 1992).

The underdetection of MDD for African Americans is of particular concern, and our findings suggested that using a cut score of 13 or 14 with African American individuals would likely result in greater false negatives due to low sensitivity. Although using a cut score of 10 or 11 maximized both sensitivity and specificity for African American participants, the overall percent correctly classified decreased substantially when using the lower cut scores. Other research has suggested that African Americans may be more inclined to endorse symptoms of psychopathology on self-report measures than to clinicians of a different ethnic background (Fresco et al., 2001). This tendency may underlie the findings of the present study, as all of our interviewers were Caucasian. Additional research should investigate this possibility further and also incorporate more individuals from other minority groups so that the accuracy of the QIDS-$SR_{16}$ as a screen for depression in ethnic minorities may be evaluated more thoroughly.

Though these results indicate that the QIDS-$SR_{16}$ is significantly better than chance at predicting a SCID-based MDD diagnosis, our findings become more meaningful in the context of the predictive power of measures commonly used to screen for MDD. A 1995 study by Mulrow and colleagues compared the utility of nine measures commonly used in screening for depression in primary care and found that sensitivities and specificities for these measures ranged from 67% to 99% and 40% to 95%, respectively. Similar values of sensitivity and specificity have been found in more recent research for depression screens commonly used in primary care settings (i.e., Bambauer et al., 2005; Beck, Steer, Ball, Ciervo, & Kabat, 1997; Wittkampf, Naeije, Schene, Huyser, & van Weert, 2007). The QIDS-$SR_{16}$ has performed comparably to these other measures in this study, with a sensitivity of 76.5% and a specificity of 81.8% using 13 or 14 as a cut score. An important area of future research is to compare the ability of the QIDS-$SR_{16}$ to that of these and other measures to correctly classify participants within the same sample as to their MDD status.

There are a number of factors that may make the QIDS-$SR_{16}$ an attractive choice as a depression screen for care providers. The QIDS-$SR_{16}$ is freely available and thus requires less monetary investment than many measures, which charge for use. The QIDS-$SR_{16}$ is time efficient in that it requires less time for training and administration relative to measures that are not designed for self-report of symptoms and/or contain a greater number of items. Also, the QIDS-$SR_{16}$ has been translated into at least 9 languages other than English, with the IDS translated into many more, and may easily be scored and interpreted by others who are not fluent in these languages. An important limitation of using the QIDS-$SR_{16}$ as a screen for MDD is that it was designed to assess depressive symptoms in the previous 7 days. This weakness could be addressed either in the context of the postassessment conversation between service provider and patient or, perhaps more simply, by altering the instructions when used as a screen to indicate that symptoms should be reported for the previous 2 weeks. To be sure, any such alterations to QIDS-$SR_{16}$ would require further validation of the measure.

The current study examined the ability of the QIDS-$SR_{16}$ to distinguish individuals with MDD from those without MDD in a community sample. Though this sample was heterogeneous in terms of psychiatric diagnosis, physical health concerns, and racial diversity, these findings may not be generalizable to other samples or other primary care settings. Future research should evaluate the QIDS-$SR_{16}$ as a screen for depression in larger and more representative samples. In our sample, participants who were retired, disabled, or unemployed were more likely to complete the SCID than those who were homemakers or employed outside of the

home. It may be that these participants had more time available to complete the SCID, either the day of their appointment or at a later time, and interest in the remuneration we offered. It may also be the case that these participants had a greater likelihood of depression due to these circumstances. For those participants who were able to complete the SCID, many were not able to complete the interview within the 3-week period, which overlapped with their QIDS-$SR_{16}$ responses. Though this time delay is certainly less than ideal, the length of the delay itself did not seem to systematically influence agreement between the QIDS-$SR_{16}$ and the SCID. Furthermore, despite the extensive training our interviewers received on administration and interpretation of the SCID for diagnostic purposes, the study may have benefited from a more thorough evaluation of interrater reliability. Finally, findings in the current study were fairly evenly divided regarding the use of 13 or 14. It would be somewhat advantageous if further research could lend more weight to the use of a single cut score for the QIDS-$SR_{16}$ in screening for MDD. In spite of these potential limitations, the results of this initial investigation indicate that the QIDS-$SR_{16}$ is an accurate and resource efficient screening measure of MDD in primary care settings.

## References


American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bambauer, K. Z., Locke, S. E., Aupont, O., Mullan, M. G., & McLaughlin, T. J. (2005). Using the Hospital Anxiety and Depression Scale to screen for depression in cardiac patients. *General Hospital Psychiatry*, *27*, 275–284.

Barkow, K., Maier, W., Üstün, T. B., Gänsicke, M., Wittchen, H. -U., & Heun, R. (2002). Risk factors for new depressive episodes in primary care: An international prospective 12-month follow-up study. *Psychological Medicine*, *32*, 595–607.

Barkow, K., Maier, W., Üstün, T. B., Gänsicke, M., Wittchen, H. -U., & Heun, R. (2003). Risk factors for depression at 12-month follow-up in adult primary health care patients with major depression: An international prospective study. *Journal of Affective Disorders*, *76*, 157–169.

Beck, A. T., Steer, R. A., Ball, R., Ciervo, C. A., & Kabat, M. (1997). Use of the Beck Anxiety and Depression Inventories for primary care with medical outpatients. *Assessment*, *4*, 211–219.

Borowsky, S. J., Rubenstein, L. V., Meredith, L. S., Camp, P., Jackson-Triche, M., & Wells, K. B. (2000). Who is at risk of nondetection of mental health problems in primary care? *Journal of General Internal Medicine*, *15*, 381–388.

Cacciola, J. S., Alterman, A. I., Rutherford, M. J., McKay, J. R., & May, D. J. (1999). Comparability of telephone and in-person structured clinical interview for DSM-III-R (SCID) diagnoses. *Assessment*, *6*, 235–242.

Corruble, E., Legrand, J. M., Zvenigorowski, H., Duret, C., & Guelfi, J. D. (1999). Concordance between self-report and clinician's assessment of depression. *Journal of Psychiatric Research*, *33*, 457–465.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders—Patient Edition (with Psychotic Screen)*. New York: New York State Psychiatric Institute.

Fresco, D. M., Coles, M. E., Heimburg, R. G., Liebowitz, M. R., Hami, S., Stein, M. B., et al. (2001). The Liebowitz Social Anxiety Scale: A comparison of the psychometric properties of self-report and clinician-administered formats. *Psychological Medicine*, *31*, 1025–1035.

Gaynes, B. N., Wisniewski, S. R., Rush, A. J., Spencer, D., Trivedi, M. H., & Fava, M. (2008). The STAR*D study: Treating depression in the real world. *Cleveland Clinic Journal of Medicine*, *75*, 57–66.

Gullion, C. M., & Rush, A. J. (1998). Toward a generalizable model of symptoms in major depressive disorder. *Biological Psychiatry*, *44*, 959–972.

Katon, W., & Ciechanowski, P. (2002). Impact of major depression on chronic medical illness. *Journal of Psychosomatic Research*, *53*, 859–863.

Katon, W., & Schulberg, H. (1992). Epidemiology of depression in primary care. *General Hospital Psychiatry*, *14*, 237–247.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 596–602.

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 617–627.

Kraemer, H. C. (1992). *Evaluating medical tests*. Newbury Park, CA: Sage.

Mulrow, C. D., Williams, J. W., Gerety, M. B., Ramirez, G., Montiel, O. M., & Kerber, C. (1995). Case-finding instruments for depression in primary care settings. *Annals of Internal Medicine*, *122*, 913–921.

Pignone, M. P., Gaynes, B. N., Rushton, J. L., Burchell, C. M., Orleans, C. T., Mulrow, C. D., et al. (2002). Screening for depression in adults: A summary of the evidence for the U.S. Preventative Service Task Force. *Annals of Internal Medicine*, *136*, 765–776.

Regier, D. A., Narrow, W. E., Rae, D. S., Manderscheid, R. W., Locke, B. Z., & Goodwin, F. K. (1993). The de facto US mental and addictive disorders service system. Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Archives of General Psychiatry*, *50*, 85–94.

Rush, A. J., Bernstein, I. H., Trivedi, M. H., Carmody, T. J., Wisniewski, S., Mundt, J. C., et al. (2006). 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.

Rush, A. J., Carmody, T., & 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.

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

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., Kilner, J., Fava, M., Wisniewski, S. R., Warden, D., Nierenberg, A. A., et al. (2008). Clinically relevant findings from STAR*D. *Psychiatric Annals*, *38*, 188–193.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., et al. (2003). 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.

StataCorp (2001). *Stata Statistical Software: Release 7.0*. College Station, TX: Author.

Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., et al. (2004). The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), 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.

Wittkampf, K. A., Naeije, L., Schene, A. H., Huyser, J., & van Weert, H. C. (2007). Diagnostic accuracy of the mood module of the Patient Health Questionnaire: A systematic review. *General Hospital Psychiatry*, *29*, 388–395.

Zanarini, M. C., Skodol, A. E., Bender, D., Dolan, R., Sanislow, C., Schaefer, E., et al. (2000). The Collaborative Longitudinal Personality Disorders Study: Reliability of Axis I and II diagnoses. *Journal of Personality Disorders*, *14*, 291–299.

RECEIVED: March 16, 2009
ACCEPTED: December 15, 2009
Available online 1 April 2010