Psychometric Properties of the Patient Health Questionnaire-9 Modified for Major Depressive Disorder in Adolescents

Aiswarya Laks Nandakumar, MBBS,¹ Jennifer L. Vande Voort, MD,¹ Paul A. Nakonezny, PhD,^{2–4} Scott S. Orth, DO,¹ Magdalena Romanowicz, MD,¹ Ayse Irem Sonmez, MD,¹ Jessica A. Ward,¹ Sandra J. Rackley, MD, MAEdHD,¹ John E. Huxsahl, MD,¹ and Paul E. Croarkin, DO, MS¹

Abstract

Objectives: The Patient Health Questionnaire-9 Modified (PHQ-9M) is a self-report tool used to assess the presence and severity of depressive symptoms in teenagers. Despite widespread use in primary care clinics and psychiatric settings, the PHQ-9M has not been validated nor are its psychometric properties adequately understood for the adolescent population. This study sought to examine the psychometrics of the PHQ-9M in treatment-seeking, depressed adolescents at a psychiatric psychopharmacology clinic who were concurrently assessed with the Children's Depression Rating Scale Revised (CDRS-R) and Quick Inventory of Depressive Symptomatology-Adolescent (17-item) Self-Report (QIDS-A17-SR).

Methods: Adolescents ($N = 160$) aged 13 through 18 years with a diagnosis of major depressive disorder, determined on the basis of a clinical interview and semi-structured interview using the Kiddie Schedule for Affective Disorders and Schizophrenia—Present and Lifetime Version, were assessed for severity of depressive symptoms with the PHQ-9M, CDRS-R (adolescent interview only), and QIDS-A17-SR assessments at baseline, 4, and 8 weeks. Classical test theory analysis was used to evaluate the internal consistency and dimensionality of the PHQ-9M. Convergent validity was evaluated via intraclass correlations of the PHQ-9M with the CDRS-R and QIDS-A17-SR. Sensitivity to treatment response was also evaluated.

Results: The internal consistency (Cronbach's coefficient α) at baseline, 4, and 8 weeks was 0.879, 0.859, and 0.827 for the PHQ-9M; 0.739, 0.835, and 0.867 for CDRS-R; and 0.712, 0.777, and 0.804 for QIDS-A17-SR, respectively. The PHQ-9M had moderate convergent validity with the CDRS-R but good convergent validity with the QIDS-A17-SR. The PHQ-9M was less sensitive to changes in symptom severity than the CDRS-R and QIDS-A17-SR.

Conclusions: The PHQ-9M appears to be a valid and reliable assessment tool for the severity of depressive symptoms in a psychiatric clinic setting. However, its utility as a treatment outcome measure may be limited compared with other available rating scales.

Keywords: adolescent, CDRS-R, depression severity, PHQ-9M, QIDS-A17-SR

Introduction

MAJOR DEPRESSIVE DISORDER (MDD) is a common medical illness worldwide and a leading cause of disability (Kessler et al. 2005; Thapar et al. 2012). The 1-year prevalence is as high as 5% by middle adolescence (Thapar et al. 2012), and the lifetime prevalence of MDD or dysthymia is as high as 11% (Merikangas et al. 2010). At least 50% of adolescents with MDD are not diagnosed until adulthood (Zuckerbrot et al. 2018), and the scarcity of psychiatric clinicians presents considerable barriers for adolescents attempting to access treatment (Erhart et al. 2009; Thapar et al. 2012; Richardson et al. 2014). Adolescents with MDD are often first seen in primary care settings (Richardson et al. 2009; Richardson et al. 2014; Zuckerbrot et al. 2018).

Evidence-based treatment approaches for MDD in adolescents include cognitive behavioral therapy (CBT), interpersonal therapy for adolescents (IPT-A), selective serotonin reuptake inhibitors (SSRIs), or the combination of psychotherapeutic approaches (CBT or IPT-A) with an SSRI (March et al. 2007; Williams et al. 2009; Wagner 2011; Cheung et al. 2013, 2018). Measurement-based care is widely viewed as the standard of care for the treatment of MDD in adolescents, although it is not always effectively applied in clinical practice (Elmquist et al. 2010; Williams et al. 2010).

A major challenge of implementing measurement-based care lies in identifying an optimal rating scale for practice settings (Williams et al. 2010; Wissow et al. 2013; Zenlea et al. 2014). An ideal rating scale would demonstrate good reliability, validity,

¹Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota.

Departments of ²Psychiatry and ³Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas.

⁴Division of Biostatistics, University of Texas Southwestern Medical Center, Dallas, Texas.

sensitivity to treatment change, and relative ease of administration (Elmqvist et al. 2010; Mayes et al. 2010). Clinical practices often identify and use one rating tool for screening, assessment, and monitoring effects of intervention (Bernstein et al. 2009; Williams et al. 2010; Cheung et al. 2018; Zuckerbrot et al. 2018).

In primary care and psychiatric clinical practice settings, the Patient Health Questionnaire-9 (PHQ-9) is being increasingly selected and used for these divergent tasks in patients across the lifespan (Richardson et al. 2010). The Children's Depression Rating Scale-Revised (CDRS-R) (Poznanski et al. 1979) and the Quick Inventory of Depressive Symptomatology-Adolescent (17-item) Self-Report (QIDS-A17-SR) (Bernstein et al. 2010) are often used to screen, assess, and monitor treatment effects during interventional studies (March et al. 2007; Mayes et al. 2010; Richardson et al. 2014).

The PHQ-9 was developed as a screening tool for depression in adults in primary care environments (Kroenke et al. 2001). The nine criteria of depression from the Diagnostic Manual of Mental Disorders, 4th edition (DSM-IV, Task Force on DSM-IV; American Psychiatric Association 1994) anchor the PHQ-9 and have severity scores ranging from 0 to 27; additional questions assess functional impairment, suicidal thoughts, and prior suicide attempts. Previous studies have shown that the PHQ-9 is a valid and sensitive screening measure for MDD in adults (Kroenke et al. 2001).

However, one study showed that the PHQ-9 had low specificity and low positive predictive value (PPV) in psychiatric specialty clinics, suggesting that it is not a valid tool for diagnostic purposes in such settings (Inoue et al. 2012). For example, patients with bipolar disorders, eating disorders, psychotic disorders, and dementia might be incorrectly identified as having a major depressive episode (Inoue et al. 2012). The wording and order of questions of the PHQ-9 were adapted for a modified version for teens (PHQ-9M) (Aggarwal et al. 2017), and additional efforts focused on streamlining of the assessments for primary care environments (Richardson et al. 2010; Haugen et al. 2016). The PHQ-9 and the modified version are now widely used for adolescents in primary care and child and adolescent psychiatry clinics but without a full understanding of their psychometric properties (Richardson et al. 2010).

Interventional studies of adolescents with depression often use other rating scales such as the CDRS-R (March et al. 2007; Emslie et al. 2008) and QIDS-A17-SR (Bernstein et al. 2010; Elmqvist et al. 2010). The CDRS-R, which was adapted from the Hamilton Depression Rating Scale, is a 17-item scale, where a raw score ≥ 40 reflects moderate symptoms of MDD and is typically the threshold score for inclusion of a patient in clinical trials for MDD (Poznanski et al. 1979). The QIDS-A17-SR is a 17-item self-report depression measure, where a score of 6–10 indicates mild depression; 10–15, moderate depression; 16–20, severe depression; and ≥ 21 , very severe depression (Bernstein et al. 2010). Both CDRS-R and QIDS-A17-SR have three components: parent, child, and interviewer impression composite scores (Poznanski et al. 1979; Bernstein et al. 2010).

Despite widespread use in primary care clinics and psychiatric settings, the PHQ-9M has not been validated in nor are its psychometric properties adequately understood for the adolescent population. Therefore, this study sought to examine the psychometric properties of the PHQ-9M compared with the CDRS-R and QIDS-A17-SR for a sample of treatment-seeking, depressed adolescents seen in a psychiatric psychopharmacology clinic. Further understanding and validation of the PHQ-9M would assist in determining its utility both for clinical practice and for research studies.

Methods

Participants

All study procedures were reviewed and approved by the Mayo Clinic Institutional Review Board before enrollment of the first participant. Participants were recruited from a child and adolescent psychopharmacology research clinic in the Mayo Clinic Depression Center, Rochester, Minnesota. The study enrolled 160 adolescents (124 girls) aged 13 through 18 years with a known history of MDD. Parents and 18-year-old patients provided informed consent, and adolescents (13–17 years) provided informed assent. Three board-certified child and adolescent psychiatrists (P.E.C., J.L.V.V., and S.S.O.) supervised all baseline clinical assessments to determine eligibility and confirm a diagnosis of MDD.

Assessments included a clinical interview and semi-structured interview using the Kiddie Schedule for Affective Disorders and Schizophrenia—Present and Lifetime Version (Kaufman et al. 1997). A baseline CDRS-R assessment was also performed; a symptom severity score of ≥ 40 was required for inclusion. Throughout the course of the study, participants were prescribed antidepressant medications, with doses adjusted on the basis of symptom severity. Patients with a history of bipolar, eating, and primary psychotic disorders were excluded. Patients with co-occurring attention-deficit/hyperactivity disorder, anxiety, post-traumatic stress disorder, and substance use (not meeting full criteria for substance abuse or use disorder) were not excluded.

Measures

Self-reported PHQ-9M scores; adolescent, parent, and composite QIDS-A17-SR scores; and adolescent, parent, and composite CDRS-R scores were collected at baseline and during the treatment course. The PHQ-9M is a self-administered assessment with nine multiple choice items, where each criterion is scored between 0 and 3, and the total score ranges from 0 to 27 (a greater score represents greater depression severity). In addition, the PHQ-9M screening assesses the patient's perception regarding the duration of depression throughout the year ("In the past year have you felt depressed or sad most days, even if you felt okay sometimes?"), functional impairment due to depressive symptoms, serious thoughts of suicide in the past month, and suicide attempts in the past year.

The QIDS-A17-SR and CDRS-R both have three components with adolescent and parent self-report scores and a composite report completed by the interviewer. The QIDS-A17-SR is a self-administered screening instrument with 17 individual criteria, each scored from 0 to 3; the total score is calculated on the basis of a scoring system where the highest score from questions 1 through 4, 5/6, 7 through 10, and 16/17 are added to scores from questions 11 through 15. The QIDS-A17-SR score ranges from 0 to 27, with a greater score representing greater depression severity.

The CDRS-R also consists of 17 items. The first 14 items of the CDRS-R are rated from the patient's or parent's responses, and the last three items are rated from the rater's observations (facial affect, speech, and hypoactivity). The CDRS-R items are rated on five-point (items 4, 5, and 16) or seven-point (items 1–3, 6–15, and 17) scales. The clinician provides an overall summary score on the basis of interviews and scores of the parent and child. Interviews with the CDRS-R are 30–40 minutes, and each total score (child rating, parent rating, and summary score) ranges from 17 to 113 (with a greater score representing greater depression severity).

Procedure

Study participants were assessed with baseline PHQ-9M self-administered screening and a total score calculated on the basis of the responses. Participants were also evaluated with baseline QIDS-A17-SR and baseline CDRS-R assessments and aggregate scores compiled. Participants completed follow-up assessments at 4- and 8-week visits.

Statistical analysis

The item and total correlation for each modified PHQ-9 item and the scale mean and standard deviation were determined by using classical test theory (CTT) analysis. Internal consistency of the PHQ-9M was determined with CTT by using Cronbach's coefficient α . Dimensionality of the PHQ-9M was defined by principal component analysis (PCA) using a varimax rotation with parallel analysis (Humphreys and Montanelli et al. 1975; Bernstein et al. 2009). With a unidimensional scale for composite scale scores from item responses, the observed eigenvalue of the first principal component should be greater than that of the first randomly generated component (eigenvalue) with parallel analysis. For all subsequent eigenvalues, the opposite effect should be observed.

Convergent validity was evaluated using the two-way random effects intraclass correlation (ICC) statistic of the PHQ-9M with criterion standard measures of depression severity (CDRS-R total and QIDS-A17-SR total). In particular, the ICC was used to assess the reliability (or consistency) between the PHQ-9M score and the CDRS-R total and QIDS-A17-SR total at baseline and at weeks 4 and 8.

The sensitivity to treatment response was evaluated by the mean percentage change in the total scores of the QIDS-A17-SR, PHQ-9M, and CDRS-R from baseline to week 4 and from baseline to week 8. Cohen's d was also calculated to estimate effect sizes for each change score.

Finally, a receiver operating characteristic (ROC) analysis was conducted along with the area under the curve (AUC) to evaluate how well the PHQ-9M measures (at baseline and at weeks 4 and 8) discriminated MDD status (CDRS-R total ≥ 40 vs. < 40). The ROC analysis also determined the optimal cut point for each measure of the PHQ-9M (based on the Youden index) in discriminating MDD status. Sensitivity and specificity in addition to PPV, negative predictive value, positive likelihood ratio (LR+), and negative LR (LR-) were also reported for the optimal cut point.

All statistical analyses were performed with SAS software, version 9.4 (SAS Institute, Inc., Cary, NC) and MedCalc for Windows, version 18.5 (MedCalc Software, Ostend, Belgium) and included CTT and PCA. The procedures in MedCalc were used to conduct the ICC and ROC analyses. The level of significance for all tests was set at an α level of 0.05 (two-tailed).

Results

Participant characteristics

The study sample included 36 adolescent boys (22.50%) and 124 girls (77.50%), with a mean age of 15.39 ± 1.47 years (range, 13–18 years). The mean age of MDD onset was 13.36 ± 2.47 years, and the mean duration of the current depressive episode was 9.33 ± 8.52 months. Mean CDRS-R total and PHQ-9M scores at baseline were 55.34 ± 10.33 and 13.31 ± 6.65 , respectively. Demographic and clinical characteristics of the 160 adolescents are summarized in Table 1.

Internal consistency and scale dimensionality

The internal consistency (Cronbach's coefficient α) of the PHQ-9M at baseline and at weeks 4 and 8 was 0.879, 0.859, and 0.827, respectively. The mean item-total correlations ranged from 0.625 at baseline to 0.573 at week 8. Table 2 summarizes the CTT results for the PHQ-9M. Cronbach's coefficient α for each of the depression scales used in this study was as follows: baseline: CDRS-R (0.739) and QIDS-A17-SR (0.712); week 4: CDRS-R (0.835) and QIDS-A17-SR (0.777); and week 8: CDRS-R (0.867) and QIDS-A17-SR (0.804).

The PCA and parallel analysis both suggested a one-factor solution [i.e., observed eigenvalues for the one principal component at baseline and at weeks 4 and 8 were 4.682, 4.390, and 4.318, respectively (Fig. 1)]. The percentage of total variance explained by the sole principal component for the PHQ-9M at baseline and at weeks 4 and 8 was 52.03%, 48.78%, 47.98%, respectively. Thus, the PCA and parallel analysis both suggested that a unidimensional-factor structure yielded the most adequate solution for the PHQ-9M (Supplementary Tables S1–S3; Supplementary Data are available online at www.liebertpub.com/cap and Fig. 1). However, examination of the factor pattern (at baseline and at weeks 4 and 8) indicated that items 3, 8, and 9 had the lowest factor loadings and communalities among the nine items that defined the unidimensional-factor structure (Supplementary Tables S1–S3).

TABLE 1. DEMOGRAPHIC AND CLINICAL CHARACTERISTICS

Characteristics	Patient sample ^{a,b}
Demographic characteristics	
Age, years	15.39 ± 1.47
Adolescent girls	124 (77.50)
White, non-Hispanic	131 (86.75) ^c
Patient characteristics	
Age of MDD onset, years	13.36 ± 2.47
Duration of current MDD episode, months	9.33 ± 8.52
MDD episodes, No.	27.94 ± 6.30
Comorbidities	
Anxiety disorder	66 (41.77) ^c
Dysthymic disorder	13 (8.33) ^c
Attention-deficit/hyperactivity disorder	28 (17.83) ^c
Substance use	9 (5.70) ^c
Clinical characteristics^d	
CDRS-R total at baseline	55.34 ± 10.33
CDRS-R total at week 4	38.48 ± 11.02
CDRS-R total at week 8	33.87 ± 11.83
QIDS-A17-SR total at baseline	15.39 ± 4.46
QIDS-A17-SR total at week 4	10.52 ± 5.40
QIDS-A17-SR total at week 8	9.21 ± 5.23
PHQ-9M at baseline	13.31 ± 6.65
PHQ-9M at week 4	9.42 ± 6.45
PHQ-9M at week 8	8.76 ± 6.27
CDRS-R total score ≥ 40 at baseline	148 (92.50)
CDRS-R total score ≥ 40 at week 4	58 (36.25)
CDRS-R total score ≥ 40 at week 8	41 (25.63)

^aValues are mean $\pm$ SD or No. (%).

^bSample size at baseline, $N = 160$.

^cPercentage based on available data.

^dCDRS-R based on an interview with the adolescent.

CDRS-R, Children's Depression Rating Scale-Revised; MDD, major depressive disorder; PHQ-9M, Patient Health Questionnaire-9 Modified; QIDS-A17-SR, Quick Inventory of Disease Symptomatology-Adolescent (17-item) Self-Report; SD, standard deviation.

TABLE 2. INTRACLAS CORRELATION COEFFICIENTS BETWEEN SCALES OF MEASUREMENT

Scale	ICC	95% CI	p Value	ICC	95% CI	p Value
<i>CDRS-R total, baseline^a</i>				<i>QIDS-A17-SR total, baseline</i>		
Baseline						
PHQ-9M (<i>n</i> = 107)	0.495	0.260–0.656	0.0001	0.729	0.602–0.805	0.0001
QIDS-A17-SR total (<i>n</i> = 158)	0.641	0.509–0.738	0.0001			
<i>CDRS-R total, week 4</i>				<i>QIDS-A17-SR total, week 4</i>		
Week 4						
PHQ-9M (<i>n</i> = 74)	0.749	0.602–0.842	0.0001	0.880	0.809–0.924	0.0001
QIDS-A17-SR total (<i>n</i> = 137)	0.746	0.644–0.819	0.0001			
<i>CDRS-R total, week 8</i>				<i>QIDS-A17-SR total, week 8</i>		
Week 8						
PHQ-9M (<i>n</i> = 62)	0.694	0.492–0.815	0.0001	0.819	0.700–0.891	0.0001
QIDS-A17-SR total (<i>n</i> = 125)	0.703	0.576–0.791	0.0001			

ICC from the two-way random effects model. ICC is the estimate of the degree of consistency or the reliability of averages of measurements.

^aCDRS-R total was based on interviews with the adolescent youth.

CDRS-R, Children's Depression Rating Scale-Revised; ICC, intraclass correlation coefficient; PHQ-9M, Patient Health Questionnaire-9 Modified; QIDS-A17-SR, Quick Inventory of Disease Symptomatology-Adolescent (17-item) Self-Report; CI, confidence interval.

Convergent validity

The mean measure ICCs between the PHQ-9M and the CDRS-R total scores at baseline and at weeks 4 and 8 were 0.495 ($p=0.0001$), 0.749 ($p=0.0001$), and 0.694 ($p=0.0001$), respectively. Moreover, the mean measure ICCs between the PHQ-9M and the QIDS-A17-SR total scores at baseline and at weeks 4 and 8 were 0.729 ($p=0.0001$), 0.880 ($p=0.0001$), and 0.819 ($p=0.0001$), respectively. The positive direction and magnitude of these ICCs suggested that the PHQ-9M had moderate consistency (or convergent validity) with the CDRS-R total but good consistency (and convergent validity) with the QIDS-A17-SR in measuring (and tracking) depression severity across the 8 weeks of treatment (Table 2).

Sensitivity to treatment response

The sensitivity to treatment response was evaluated by the mean percentage change in total scores of the PHQ-9M, QIDS-A17-SR, and CDRS-R from baseline to week 4 and from baseline to week 8.

The mean percentage change (decrease) in the total scores from baseline to week 4 was -17.79% (95% confidence interval [CI]: -32.90 to -2.68 , $p=0.02$; Cohen's $d=0.302$) in the PHQ-9M; -29.32% (95% CI: -35.09 to -23.56% , $p<0.0001$; Cohen's $d=0.925$) in the QIDS-A17-SR; and -28.56% (95% CI: -31.99 to -25.17 , $p<0.0001$; Cohen's $d=1.622$) in the CDRS-R.

Moreover, the mean percentage change (decrease) in the total scores from baseline to week 8 was -2.53% (95% CI: -41.74 to 36.67 , $p=0.89$; Cohen's $d=0.022$) in the PHQ-9M, -36.47% (95% CI: -42.47 to -30.48 , $p<0.0001$; Cohen's $d=1.228$) in the QIDS-A17-SR; and -36.89% (95% CI: -40.68 to -33.11 , $p<0.0001$; Cohen's $d=1.922$) in the CDRS-R. These results suggested that the PHQ-9M was less sensitive to changes in symptom severity than both the QIDS-A17-SR and the CDRS-R.

ROC results

The ROC analysis (based on the Youden index) showed that the PHQ-9M with a cutoff >6 (baseline), >8 (week 4), and >7 (week 8) best discriminated MDD status (CDRS-R total ≥ 40 vs. <40) with

FIG. 1. Scree plots for the Patient Health Questionnaire-9 Modified. (A), Baseline, (B) week 4, (C) week 8.

a reasonable degree of accuracy (AUC: 0.72 at baseline, 0.89 at week 4, and 0.75 at week 8) and with good PPV at baseline (PPV=97.80%) and at week 4 (PPV=77.80%) but poor PPV at week 8 (47.10%). We note that the LR+ of 2.58 (baseline), 4.85 (week 4), and 2.06 (week 8) indicated an increased likelihood (or probability) that the adolescent with MDD had a CDRS-R total score ≥ 40 on the basis of the aforementioned PHQ-9M cutoffs (>6 , >8 , >7 , respectively). The ROC results are presented in Table 3.

Discussion

To our knowledge, this was the first study to validate and examine the psychometric properties of the PHQ-9M as a depression symptom-severity rating tool for treatment-seeking adolescents. The results are encouraging: the PHQ-9M had moderate convergent validity with the CDRS-R and good convergent validity with the QIDS-A17-SR for the measurement of depressive symptom severity across 8 weeks of treatment. ROC analyses suggested that the PHQ-9M discriminated MDD status (on the basis of CDRS-R score) with a reasonable degree of accuracy and PPV.

Given its ease of use, the PHQ-9M appears to be an acceptable choice for assessing depressive symptoms in an adolescent psychiatric clinic setting. The QIDS-A17-SR may have a greater time burden for patients and providers but otherwise appears to be a superior choice to the PHQ-9M. Notably, the PHQ-9M was less sensitive to changes over 8 weeks of treatment in depressive symptom severity than the CDRS-R and QIDS-A17-SR. On the basis of the present findings, the PHQ-9M is not an optimal measure for interventional research or for tracking symptom improvement over the course of treatment. In a clinical setting, the QIDS-A-17-SR would have more utility compared with CDRS-R. Notably, the CDRS-R is not in the public domain and requires a trained clinical interviewer.

The QIDS-A17-SR was more sensitive to change as compared to the PHQ-9M. This finding was somewhat unexpected as the PHQ-9 and QIDS-A17-SR had moderate agreement (see ICC in Table 2) at baseline (ICC=0.729). The magnitude of the agreement between the PHQ-9M and the QIDS-A17-SR increased at week 4 (ICC=0.88) but then decreased in magnitude from weeks 4 to 8 (ICC=0.819). However, when examining sensitivity to treatment response

evaluated by the mean percentage changes in rating scales, the QIDS-A17-SR was more sensitive to change. From a clinical perspective, the QIDS-A17-SR has more items compared with the PHQ-9M and arguably collects broader information regarding depressive symptom severity.

Strengths of the present study include a well-characterized sample of depressed adolescents, face-to-face structured clinical interviews, and use of standard psychiatric research tools. The utility, validity, and reliability of the PHQ-9M in child and adolescent psychiatry specialty clinics is inadequately understood (Elmqvist et al. 2010; Richardson et al. 2010; Aggarwal et al. 2017), but the current study addresses these concerns.

A prior study by Richardson et al. (2010) showed that the PHQ-9 had good sensitivity and specificity for the recognition of MDD in adolescents in primary care environments. Although their study had a number of important strengths, it examined the standard PHQ-9. Further, data were collected through surveys and telephone interviews (Richardson et al. 2010). Although the use of telephone interviews and surveys is widespread in health services research, the face validity of this approach for diagnosing and assessing adolescent depression disorders is questionable (Rohde et al. 1997; Erhart et al. 2009). One study has described the variable reliability among telephone interviews, surveys, and direct interviews (Erhart et al. 2009).

The United States Preventive Services Task Force has recommended screening for MDD in adolescents aged 12 to 18 years when adequate systems are in place to ensure accurate diagnosis, effective treatment, and appropriate follow-up (US Preventive Services Task Force 2009). The Guidelines for Adolescent Depression in Primary Care recently reinforced this need by recommending implementation of annual universal screening of youth 12 and over at health maintenance visits to effectively tackle the greater clinical need and shortage of mental health specialists (Cheung et al. 2018; Zuckerbrot et al. 2018).

Although depression screening in primary care settings is recognized as an essential measure to identify patients at risk, the implementation of screening with rating tools in community-based physician practices is inconsistent (Cheung et al. 2010, 2013, 2018; Zuckerbrot et al. 2018). This may be related to the ambiguity regarding optimal screening tools (Elmqvist et al.

TABLE 3. RESULTS OF RECEIVER OPERATING CHARACTERISTICS FOR THE PHQ-9M AND BINARY MEASUREMENT OF CDRS-R TOTAL AT BASELINE AND AT WEEKS 4 AND 8

PHQ-9M scale	CDRS-R total (≥ 40 vs. < 40) ^a								95% CI AUC ^d
	Cut point ^b	Sensitivity	Specificity	PPV, %	NPV, %	LR+	LR–	AUC (SE) ^c	
Baseline									
PHQ-9M ($n = 107$)	>6.00	86.14	66.67	97.80	22.20	2.58	0.21	0.72 (0.12) ^e	0.62–0.80
Week 4									
PHQ-9M ($n = 74$)	>8.00	90.32	81.40	77.80	92.10	4.85	0.12	0.89 (0.04) ^f	0.80–0.95
Week 8									
PHQ-9M ($n = 63$)	>7.00	84.21	59.09	47.10	89.70	2.06	0.27	0.75 (0.06) ^f	0.62–0.85

^aCDRS-R total, baseline: ≥ 40 , $n = 101$; < 40 , $n = 6$. CDRS-R total, week 4: ≥ 40 , $n = 31$; < 40 , $n = 43$. CDRS-R total, week 8: ≥ 40 , $n = 19$; < 40 , $n = 44$.

^bCut point for PHQ-9M was based on the Youden index.

^cDeLong et al. method was used for calculation of the standard error of the AUC.

^dBinomial exact CI for the AUC.

^e $p = 0.08$.

^f $p < 0.0001$.

AUC, area under the curve; CDRS-R, Children's Depression Rating Scale-Revised; LR+, positive likelihood ratio; LR–, negative likelihood ratio; NPV, negative predictive value; PHQ-9M, Patient Health Questionnaire-9 Modified; PPV, positive predictive value; SE, standard error; CI, confidence interval.

2010) and the variability of implementing screening in child and adolescent psychiatric specialty clinics. Having unified approaches in community-based primary care, tertiary care, and research settings for screening and assessing depressive symptoms in adolescents would optimize care, assist with forward and reverse research translations in psychiatry, and mitigate rising health care costs (Elmqvist et al. 2010; Roseman et al. 2016). One study has examined the use of the PHQ-9M in rural, semirural, and impoverished areas (Aggarwal et al. 2017); however, the PHQ-9M has not been adequately studied in research settings.

Prior studies have recommended an optimal cut-point score of ≥ 11 on the PHQ-9 for the detection of MDD in adolescents compared with a score of ≥ 10 in the adult population (Richardson et al. 2010). Our findings suggest that lower cut-point scores on the PHQ-9M might be considered for the diagnosis of MDD or recognition of moderate symptoms of MDD in adolescents evaluated or treated in a specialty clinic. The present findings suggest that cut-point scores of 6–8 on the PHQ-9M (depending upon the temporal assessment period) may warrant further study as threshold values of MDD in child and adolescent psychiatric specialty clinics.

Interventional research studies focused on depression in adolescents have considerable challenges such as data quality, identification of participants, mitigating placebo response, and demonstrating change in symptom severity (Walkup 2017; Strawn and Croarkin 2018). An ideal outcome measure for interventional research for adolescent MDD would demonstrate sensitivity to change in symptom severity (Bernstein et al. 2010; Elmqvist et al. 2010; Mayes et al. 2010). Arguably, optimal sensitivity to change is correspondingly important for integrating rating scales in clinical practice.

Unfortunately, our current findings suggest that the PHQ-9M may have limited sensitivity to change in symptom severity in adolescents with MDD treated in a specialty clinic. While the CDRS-R was superior with respect to sensitivity to change in symptom severity, the associated costs and time burden often create challenges for research studies and render its use unfeasible in primary care environments. Conversely, the QIDS-A17-SR was superior to the PHQ-9M with respect to sensitivity to change and may have a similar time burden (as the CDRS-R) for patients and clinicians when compared with the PHQ-9M.

This study has limitations. First, the sample was of moderate size for a psychometric study and was also predominantly composed of adolescent girls. The sample included only treatment-seeking adolescents in a specialty psychiatric clinic, thus limiting the generalizability of findings. Finally, this study examined a convenience sample. Notwithstanding these limitations, this study addresses the paucity of studies in the literature and advances the collective understanding of the psychometric properties of the PHQ-9M for adolescents with MDD in a psychiatric specialty clinic. Ideally, further work will examine the possibility of a cohesive rating scale for various clinical and research settings.

Conclusions

In summary, this study examined the psychometric properties of the PHQ-9M in a psychiatric clinic compared with the QIDS-A17-SR and the CDRS-R. The PHQ-9M appears to have good internal consistency and convergent validity.

Clinical Significance

The PHQ-9M appears to perform relatively well as a screening tool for MDD in a specialty environment. Further study of lower

PHQ-9M cut-point scores for identification of MDD in specialty clinics may have utility. Unfortunately, the PHQ-9M may not be an ideal measure for monitoring the temporal effects of treatment or for assessing the sensitivity to change in symptom severity.

Acknowledgment

Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award K23MH100266. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

A.L.N., J.L.V.V., S.S.O., M.R., A.I.S., S.J.R., J.E.H., and P.E.C. contributed to the design of the study and interpretation of data. A.L.N., J.L.V.V., S.S.O., J.A.W., and P.E.C. acquired the data. P.A.N. completed statistical analyses. A.L.N., J.L.V.V., P.A.N., S.S.O., M.R., A.I.S., J.A.W., S.J.R., J.E.H., and P.E.C. drafted, revised, critically reviewed, and approved the final submitted draft of the article. P.E.C., A.L.N., and P.A.N. had full access to all of the data in the study. P.E.C. takes responsibility for the integrity of the data and the data analysis.

Disclosures

P.E.C. has received research grant support from Pfizer, Inc., NIMH (K23 MH100266 and R01MH113700), the Brain and Behavior Research Foundation, and the Mayo Clinic Foundation. He has served as a site subprincipal or principal investigator (without additional compensation) for Eli Lilly and Co, Forest Laboratories, Inc., Merck & Co, Inc., and Pfizer, Inc.; has received equipment support from Neuronetics, Inc.; and receives supplies and genotyping services from Assurex Health, Inc., for an investigator-initiated study. He is the primary investigator for a multicenter study funded by Neuronetics, Inc. He is a site primary investigator for a study funded by NeoSync, Inc. All other authors have no disclosures.

References

- Aggarwal S, Taljard L, Wilson Z, Berk M: Evaluation of modified patient health questionnaire-9 teen in South African adolescents. *Indian J Psychol Med* 39:143–145, 2017.
- American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, DC: American Psychiatric Association; 1994.
- Bernstein IH, Rush AJ, Trivedi MH, Hughes CW, Macleod L, Witte BP, Jain S, Mayes TL, Emslie GJ: Psychometric properties of the Quick Inventory of Depressive Symptomatology in adolescents. *Int J Methods Psychiatr Res* 19:185–194, 2010.
- Bernstein IH, Wendt B, Nasr SJ, Rush AJ: Screening for major depression in private practice. *J Psychiatr Pract* 15:87–94, 2009.
- Cheung A, Mayes T, Levitt A, Schaffer A, Michalak E, Kiss A, Emslie G: Anxiety as a predictor of treatment outcome in children and adolescents with depression. *J Child Adolesc Psychopharmacol* 20:211–216, 2010.
- Cheung AH, Kozloff N, Sacks, D: Pediatric depression: An evidence-based update on treatment interventions. *Curr Psychiatry Rep* 15: 381, 2013.
- Cheung AH, Zuckerbrot RA, Jensen PS, Laraque D, Stein REK, Glad-Pc Steering Group: *Guidelines for Adolescent Depression in Primary Care (GLAD-PC): Part II. Treatment and ongoing management*. *Pediatrics* 120:e1313, 2018.

- DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837–845, 1988.
- Elmqvist JM, Melton TK, Croarkin P, McClintock SM: A systematic overview of measurement-based care in the treatment of childhood and adolescent depression. *J Psychiatr Pract* 16:217–234, 2010.
- Emslie GJ, Kennard BD, Mayes TL, Nightingale-Teresi J, Carmody T, Hughes CW, Rush AJ, Tao R, Rintelmann JW: Fluoxetine versus placebo in preventing relapse of major depression in children and adolescents. *Am J Psychiatry* 165:459–467, 2008.
- Erhart M, Wetzel RM, Krugel A, Ravens-Sieberger U: Effects of phone versus mail survey methods on the measurement of health-related quality of life and emotional and behavioural problems in adolescents. *BMC Public Health* 9:491, 2009.
- Haugen W, Haavet OR, Sirpal MK, Christensen KS: Identifying depression among adolescents using three key questions: A validation study in primary care. *Br J Gen Pract* 66:e65–e70, 2016.
- Humphreys LG, Montanelli RG, Jr.: An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behav Res* 10:193–205, 1975.
- Inoue T, Tanaka T, Nakagawa S, Nakato Y, Kameyama R, Boku S, Toda H, Kurita T, Koyama T: Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care. *BMC Psychiatry* 12: 73, 2012.
- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N: Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 36:980–988, 1997.
- Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE: Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 62:617–627, 2005.
- Kroenke K, Spitzer RL, Williams JB: The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med* 16:606–613, 2001.
- March JS, Silva S, Petrycki S, Curry J, Wells K, Fairbank J, Burns B, Domino M, McNulty S, Vitiello B, Severe J: The treatment for adolescents with depression study (TADS): Long-term effectiveness and safety outcomes. *Arch Gen Psychiatry* 64:1132–1143, 2007.
- Mayes TL, Bernstein IH, Haley CL, Kennard BD, Emslie GJ: Psychometric properties of the Children's Depression Rating Scale-revised in adolescents. *J Child Adolesc Psychopharmacol* 20:513–516, 2010.
- Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, Benjet C, Georgiades K, Swendsen J: Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication—Adolescent Supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry* 49:980–989, 2010.
- Poznanski EO, Cook SC, Carroll BJ: A depression rating scale for children. *Pediatrics* 64:442–450, 1979.
- Richardson L, McCauley E, Katon W: Collaborative care for adolescent depression: A pilot study. *Gen Hosp Psychiatry* 31:36–45, 2009.
- Richardson LP, Ludman E, McCauley E, Lindenbaum J, Larison C, Zhou C, Clarke G, Brent D, Katon W: Collaborative care for adolescents with depression in primary care: A randomized clinical trial. *JAMA* 312:809–816, 2014.
- Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, Rockhill C, Katon W: Evaluation of the Patient Health Questionnaire-9 item for detecting major depression among adolescents. *Pediatrics* 126:1117–1123, 2010.
- Rohde P, Lewinsohn PM, Seeley JR: Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. *Am J Psychiatry* 154:1593–1598, 1997.
- Roseman M, Kloda LA, Saadat N, Riehm KE, Ickowicz A, Baltzer F, Katz LY, Patten SB, Rousseau C, Thombs BD: Accuracy of depression screening tools to detect major depression in children and adolescents: A systematic review. *Can J Psychiatry* 61:746–757, 2016.
- Strawn JR, Croarkin PE: Desvenlafaxine versus placebo in a fluoxetine-referenced study of children and adolescents with major depressive disorder: Design, definitions, and ongoing challenges for child and adolescent psychopharmacology research. *J Child Adolesc Psychopharmacol* 28:363, 2018.
- Thapar A, Collishaw S, Pine DS, Thapar AK: Depression in adolescence. *Lancet* 379:1056–1067, 2012.
- US Preventive Services Task Force: Screening and treatment for major depressive disorder in children and adolescents: US Preventive Services Task Force recommendation statement. *Pediatrics* 123:1223–1228, 2009.
- Wagner KD: Pediatric and adolescent depression. *Pediatr Ann* 40: 290–291, 2011.
- Walkup JT: Antidepressant efficacy for depression in children and adolescents: industry- and NIMH-funded studies. *Am J Psychiatry* 174:430–437, 2017.
- Williams MD, Jaekels N, Rummans TA, Somers K, Nesse RE, Gorman RS: Creating value in depression management. *Qual Prim Care* 18:327–333, 2010.
- Williams SB, O'Connor EA, Eder M, Whitlock EP: Screening for child and adolescent depression in primary care settings: A systematic evidence review for the US Preventive Services Task Force. *Pediatrics* 123:e716–e735, 2009.
- Wissow LS, Brown J, Fothergill KE, Gadowski A, Hacker K, Salmon P, Zerkowicz R: Universal mental health screening in pediatric primary care: A systematic review. *J Am Acad Child Adolesc Psychiatry* 52:1134–1147.e23, 2013.
- Zenlea IS, Milliren CE, Mednick L, Rhodes ET: Depression screening in adolescents in the United States: A national study of ambulatory office-based practice. *Acad Pediatr* 14:186–191, 2014.
- Zuckerbrot RA, Cheung A, Jensen PS, Stein REK, Laraque D, Glad-Pc Steering Group: Guidelines for Adolescent Depression in Primary Care (GLAD-PC): Part I. Practice preparation, identification, assessment, and initial management. *Pediatrics* 2018 [Epub ahead of print]; DOI: 10.1542/peds.2017-4081.

Address correspondence to:

Paul E. Croarkin, DO, MS

Department of Psychiatry and Psychology

Mayo Clinic

200 First Street SW

Rochester, MN 55905

E-mail: croarkin.paul@mayo.edu