Evaluation of the Patient Health Questionnaire (PHQ-9) for Detecting Major Depression among Adolescents

Laura P. Richardson, MD, MPH[1,4,3], **Elizabeth McCauley, PhD**[2,4], **David C. Grossman, MD, MPH**[3,5,1], **Carolyn A. McCarty, PhD**[1,4], **Julie Richards, MPH**[3], **Joan E. Russo, PhD**[2], **Carol Rockhill, MD, MPH**[2,4], and **Wayne Katon, MD**[2]

[1]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA

[2]Department of Psychiatry & Behavioral Sciences, University of Washington School of Medicine, Seattle, WA

[3]Group Health Research Institute, Seattle, WA

[4]Seattle Children's Hospital, Seattle, WA

[5]Department of Health Services, University of Washington School of Public Health

Abstract

Objective—The purpose of this study was to examine the performance characteristics and validity of the Patient Health Questionnaire - 9 item (PHQ-9) as a screening tool for depression among adolescents.

Methods—The PHQ-9 was completed by 442 youth (13-17 years) who were enrolled in a large healthcare delivery system and participating in a study on depression outcomes. Criterion validity and performance characteristics were assessed against an independent structured mental health interview (the Child Diagnostic Interview Schedule, DISC-IV). Construct validity was tested by examining associations between the PHQ-9 and a self-report measure of functional impairment, as well as parental reports of child psychosocial impairment and internalizing symptoms.

Results—A PHQ-9 score ≥11 had a sensitivity of 89.5% and specificity of 77.5% for detecting youth meeting DSM-IV criteria for major depression on the DISC-IV. On ROC analysis the PHQ-9 had an area under the curve of 0.88 (95% CI = 0.82 to 0.94) and the cut point of 11 was optimal for maximizing sensitivity without loss of specificity. Increasing PHQ-9 scores were significantly correlated with increasing levels of functional impairment, as well as parental report of internalizing symptoms and psychosocial problems.

Conclusion—Although the optimal cut point is higher among adolescents, the sensitivity and specificity of the PHQ-9 are similar to those of adult populations. The brief nature and ease of scoring of this instrument make this tool an excellent choice for providers and researchers seeking to implement depression screening in primary care settings.

Keywords

depression; adolescents; screening; primary care

Corresponding Author: Laura Richardson, MD, MPH Center for Child, Health, Behavior and Development Seattle Children's Hospital Research Institute 1100 Olive Way, Suite 500; M/S MPW 8-1 Seattle, WA 98101 PH: 206 884-8245 FAX: 206-884-7801 laura.richardson@seattlechildrens.org.

Introduction

In response to the growing evidence for effective treatments for depression among adolescents, the US Preventive Services Task Force now recommends screening for depression among adolescents in primary care settings.[1] However, a recent systematic review identified only five studies with adequate psychometric data on sensitivity and specificity of depression screening instruments in primary care.[2] Two of the studies evaluated the same instrument, thus four screening instruments were tested. The four instruments identified each have potential limitations for application in primary care including cost to administer a patented instrument,[3] length of the screening instrument,[4, 5] algorithm based scoring rather than symptom scoring,[6] and the ability to be a stand alone screener for depression.[7]

Ideal screening instruments are brief, easy to understand for patients, simple to score, available without cost, and have strong performance characteristics. The Patient Health Questionnaire 9-item (PHQ-9) depression screener was developed for administration among adults in primary care settings. It has been shown to have good diagnostic validity and comparable sensitivity and specificity to other longer measures of depression.[8, 9] In addition, because the instrument is based on DSM-IV criteria, the same 9-items are used in adults to establish probable depressive disorder diagnoses as well as to grade depressive symptom severity.

Despite wide use in adult populations, the PHQ-9 has not been validated in adolescent populations. In this paper, we evaluate the operating characteristics (sensitivity and specificity) of the PHQ-9 as a diagnostic instrument for depressive disorders among adolescents and the construct validity of the PHQ-9 as a depression severity measure in relation to functional status and parental assessment of child symptoms.

Methods

The AdoleSCent Health (ASC) Study was developed by a multidisciplinary team at the University of Washington and the Group Health (GH) Research Institute. The main purposes of ASC Study were to evaluate the performance of depression screening tools and to describe the characteristics of adolescents who would most benefit from exposure to evidence-based interventions for depression in primary care settings. All procedures were approved by the GH Human Subjects Protection Committee.

Survey Methods

Between September 2007 and June 2008, study staff randomly selected 4,000 enrollees, ages 13-17, who had seen a provider in a GH healthcare facility at least once in the prior 12 months. GH is a consumer-governed non-profit healthcare organization that serves over 630,000 residents of Washington and Idaho. The parents/guardians of selected enrollees received an invitation letter, a consent form, and a brief survey for their child. Parents were asked to sign the consent form and give the survey to their child to complete in a private place. The child received a $2 pre-incentive. Completion of the survey was taken as a form of assent by the child. Parents of youth who did not respond received a second mailing and follow-up phone calls.

The brief survey consisted of 10-items about age, gender, weight, height, sedentary behaviors, overall heath, functional impairment, and depressive symptoms (the Patient Health Questionnaire 2-item Depression Scale (PHQ-2)). The PHQ-2 includes the first two items from the PHQ-9 and asks respondents to rate the frequency (0 = not at all; 3 = nearly every day) that they have had: a) depressed mood and/or b) lack of pleasure in usual

activities in the past 2 weeks. In a prior publication using these same data, we found that a score of ≥3 on the PHQ-2 has a sensitivity of 73% and a specificity of 75% for detecting major depression among adolescents.[10]

A subset of youth (n = 499) were invited to participate in the follow-up phone interview study, during which in-depth information was obtained on depressive symptoms, functional impairment, and health behaviors. Youth with a PHQ-2 ≥ 3 (n = 271) and a sample of youth with a PHQ-2 ≤ 2 (n = 228) who were frequency matched for age and gender were invited to participate. Youth completing the follow-up interview received $20. Consent for the phone survey was obtained from both the parent and the child.

Depression measures

The child phone interview included the Patient Health Questionnaire 9-item (PHQ-9) screener and the Diagnostic Interview Schedule for Children depression modules (DISC-IV). The PHQ-9 was completed prior to other depression and mental health measures.

The PHQ-9 is a self-administered version of the depression portion of the PRIME-MD interview,[11] which uses DSM-IV criteria to assess for mental disorders in primary care.[8] It can be scored to provide a dichotomous diagnosis of probable major depression and to grade symptom severity via a continuous score. The PHQ-9 has been found to have high sensitivity (73%) and high specificity (98%) for the diagnosis of major depression in adult populations.[8, 11] Among adults, scores on the PHQ-9 have also been used to define severity for probable diagnoses in the following manner: a score of 5-9 is considered minimal depression, 10-14 is considered mild major, 15-19 is moderate major, and ≥20 is severe major.[8] The PHQ-9 also has a functional impairment question (item 10) that asks how much the symptoms they endorse in the first 9 items interfere with daily functioning.

The DISC-IV is a reliable and valid structured interview designed for lay interviewers, which includes algorithms to diagnose DSM-IV disorders in children and adolescents.[12] Telephone versions of structured psychiatric interviews have been found to have a high correlation with in-person interviews.[13, 14] In order to decrease patient burden, only the depression modules (major depression and dysthymia) were used. All interviewers received 12 hours of classroom and hands-on training and additional project-specific training on the DISC-IV.

Other Measures

The Columbia Impairment Scale (CIS) was used to assess functional impairment.[15] The 13-item CIS scale measures adolescent impairment in school, family, and peer relationships and has been shown to correlate with the clinician-rated Children's Global Assessment Scale.[15]

To assess for anxiety symptoms, youth were asked to complete the brief 5-item version of the Screen for Child Anxiety Related Emotional Disorders (SCARED).[16] Using a cut-off of 3 or greater, the brief SCARED has been shown to have a sensitivity of 74% and a specificity of 73% for discriminating youth with clinically significant anxiety from those without anxiety.[16]

Parents were asked to complete the Brief Pediatric Symptom Checklist (PSC-17). The internalizing component of the PSC-17 (at a cut-point of ≥5) has a sensitivity of 73% and specificity of 74% for detecting youth with a depressive disorder.[17] The externalizing component (at a cut point of ≥7) has a sensitivity of 62% and a specificity of 89% for detecting youth who met criteria for an externalizing disorder.[17]

Statistical Analysis

Descriptive statistics were completed for the full sample and stratified by depression status. Three categories of depression status were used based on algorithms from the DISC-IV: major depression, "intermediate depression", or no depressive disorder.[12] Youth with "intermediate depression" reported at least 3 of the 9 symptom criteria for major depression but did not meet diagnostic criteria for major depression in the past year. Chi-square analyses and F-test analyses were used to compare categorical and continuous variables, respectively, among depressed and non-depressed individuals based on PHQ-9 score. The area under the ROC curve was calculated as a quantification of the sensitivity and specificity of the ability of the self-report questionnaires to classify youth into the past month major depression category based on the DISC-IV. Results were interpreted based on standards that have been set for interpreting the area under the curve.[18]

Results

Of 3,775 eligible youth, 2291 (60.7%) completed the brief survey (Figure 1). Twelve percent of youth (N=281) screened positive for possible depression with a PHQ-2 ≥ 3. 499 youth were invited to participate in the full baseline assessment of whom 444 (89%) consented and both the parent and child completed the baseline survey. Two youth who met DSM-IV criteria for bereavement on the DISC-IV were removed from the analytic sample resulting in a final sample of 442 youth.

The mean age of participants was 15.3 years (SD = 1.2) and 60% of subjects were female. The sample was predominantly White (71%) with the next largest minority groups being Asian (10%) and Black (9.6%). Seventy-eight percent came from two parent households and 87% of youth came from households where at least one parent had at least some college. The median neighborhood household income for participants was $57,442 (SD = $18,293). Among the 242 youth who had a positive PHQ-2 on initial screening, 112 were still positive on a PHQ-2 two weeks later and 101 had a PHQ-9 score of 11 or higher. Among the 202 that were negative on the screening PHQ-2, 194 were still negative on the PHQ-2 two weeks later and 190 had a PHQ-9 score less than 11.

Table 1 shows the distribution of PHQ-9 scores by depression status on the DISC-IV. Categories on the PHQ-9 (minimal, mild, moderate, moderately severe, & severe) were based on severity thresholds established by the original authors of the PHQ-9 [8] with the exception of the use of 11 rather than 10 as the lower threshold for the moderate category based on the results of our ROC analyses. Youth meeting criteria for major depression had significantly higher total scores on the PHQ-9 and were more likely to be in the "moderate" to "severe" categories of impairment on the PHQ-9 compared to the other two diagnostic groups (Chi-square (8) = 105.06, p < .0001). Youth with "intermediate depression" on the DISC-IV were most likely to be in the "mild" to "moderate" categories of impairment on the PHQ-9. Youth with no depression diagnosis were most likely to report minimal symptoms. The mean PHQ-9 scores also decreased significantly and linearly from a high of 15.5 (SD = 5.6) for those with major depressive disorder, 10.7 (SD = 3.9) for those with "intermediate depression", and 6.1 (SD = 5.1) for youth with no depressive disorder (F(2,439) = 48.08, p < .0001).

Table 2 shows the test characteristics of the PHQ-9 using the DISC-IV as a gold standard. The optimal cut-point for maximizing sensitivity of the PHQ-9 without loss of specificity was a score of 11 or greater. At this cut-point, the PHQ-9 had a sensitivity of 89.5% and a specificity of 77.5% for detecting youth with major depression on the DISC-IV. The positive predictive value was 15.2% for detecting major depression on the DISC-IV and the negative

predictive value was 99.4%. On ROC analysis (Figure 2) the area under the curve for detecting major depression was 0.88 (0.82 – 0.94).

We also assessed the PHQ-9 in our sample using the algorithmic scoring protocol for probable major depression (presence of depression and/or anhedonia at least "more than half the days" and a minimum of 5 total symptoms occurring at least "more than half the days" (with the exception of suicide which is positive with any endorsement)). Using this scoring protocol, the sensitivity for detecting youth with Major Depression on the DISC-IV was 57.9% and the specificity was 90.3%.

Table 3 shows the relationship between PHQ-9 scores and each of our measures of impairment. Scores on the Columbia Impairment Scale (CIS), mean depressive symptom related difficulty (item 10 on the PHQ-9), parental report of internalizing symptoms (PSC-17 internalizing scale), and overall psychosocial impairment (total PSC-17 score on the parent version) increased in a linear fashion such that youth with higher PHQ-9 scores also exhibited higher scores (indicating more impairment) on each of these measures (p <0.0001 for all measures).

The false positive rate, calculated as 1-specificity, was approximately 22.5% when using the DISC-IV as a gold standard. To better understand the characteristics of youth with false positives, we examined the association between having a false positive PHQ-9 and having "intermediate depression" or a positive screening test for another disorder (anxiety or externalizing disorder). Among adolescents with PHQ-9 scores ≥11 but no DISC-IV diagnosis for major depression (N = 95), 29.3% had "intermediate depression" on the DISC-IV, 16.5% had major depression in the past year but not in the prior month, 23.9% had elevated externalizing disorder symptoms (PSC-17 externalizing scale score ≥7), and 56.8% had clinically significant anxiety symptoms (SCARED score ≥3). Taken together, 82.2% of the false positive group had at least one of the four indications we examined: 45.3% had one, 31.6% had two, and 5.3% had three of the four indications.

Discussion

The US Preventive Services Task Force advises primary clinicians to screen adolescents for depression provided there is a system of care to confirm diagnosis and initiate treatment.[1] To implement this recommendation, providers need screening tools that can be easily implemented in pediatric primary care settings. Although it has been extensively tested among adults, this is the first study to examine the test characteristics of the PHQ-9 in an adolescent population. We found that, when compared to a structured diagnostic interview, the PHQ-9 had high sensitivity (89.5%) and good specificity (78.8%) for detecting major depression among adolescents and on ROC analysis had an area under the curve of 0.88 putting this screening tool in a "good" range. [18] This sensitivity and specificity of the PHQ-9 is in a similar range to other depression screening tools that have been tested among adolescents in primary care: BDI (Sensitivity – 91%, Specificity – 91%)[3], PHQ-A (Sensitivity – 73%, Specificity – 94%)[6], and Short MFQ[19] (Sensitivity – 80%, Specificity – 81%)[20], and performs better than physician interview following targeted training (Sensitivity – 43% and Specificity – 87%).[5]

In adult samples, a PHQ-9 score of 10 or higher is recommended to identify individuals with likely depression. Based on our findings, we would recommend using a cut-point of 11 or higher to indicate the need for further evaluation for depression. However, providers may reasonably choose an alternate cut point. For example, even though it may result in a higher rate of false positives, clinics where both adolescents and adults are seen might choose to use a cut-off of 10 in order to simplify procedures for providers.

In a prior study using this sample, we found that the PHQ-2 which contains the first two items of the PHQ-9 has a sensitivity of 74% and a specificity of 75% for detecting major depression among adolescents.[10] Clinics wishing to minimize respondent burden could start with the PHQ-2 followed by the full PHQ-9 only on those with a score of 3 or higher on the PHQ-2. The benefit of adding the PHQ-9 in this protocol is that it provides more information on individual depressive symptoms, has better specificity for major depression than the PHQ-2, and it includes a question about suicide, an important cause of mortality among adolescents.[21] It is important to note, however, in using the PHQ-9 that youth do not need to be depressed to be suicidal. Any positive indication of suicidality (a score of 1 or higher on item 9 of the PHQ-9) should be taken seriously and followed up on by providers regardless of total PHQ-9 score.

Compared to the findings in adults, the sensitivity of the PHQ-9 is higher but the specificity is lower in the adolescent population. This suggests that when used as a screening tool, the PHQ-9 is less likely to miss youth with major depression but there is a higher false positive rate in adolescent populations. The higher false positive rate in adolescent populations may be a result of a high rate of subthreshold depressive symptoms and adjustment disorders, as well as a significant overlap of symptoms between mental health disorders among this age group. Of the youth who were in the false positive category, 82% had an indication of a mental health concern including meeting criteria for "intermediate depression" on the DISC-IV, having depression in the past year but not in the past month, having high levels of externalizing behavior and/or having high levels of anxiety symptoms suggesting the need for further monitoring.

An additional difference between the adult and the youth DSM-IV criteria for major depressive disorder is that youth may meet the diagnostic criteria by presenting with irritability rather than depressed mood. The PHQ-9 does not include an item about irritability and, to allow for the use of a single form for settings where both adolescents and adults are seen, we chose not to change the wording of the PHQ-9. As we did not add an irritability item, we are not able to determine how it may have modified the performance of the PHQ-9. The DISC-IV does include an irritability item and some of the discrepancy between these two instruments may relate to this difference.

This study has the following limitations. First, this study was conducted in an insured population of adolescents in the Pacific Northwest and may not be generalizable to all adolescent populations. Second, the response rate to our initial brief screen was 60% and we may have had some selection bias regarding youth who participated in the study. Although we were very encouraged by the 89% participation rate in the follow-up interview study, it is possible that youth who chose not to participate were different from those who did. Third, since we oversampled youth with elevated PHQ-2 scores, the prevalence of depression in our study sample may be higher than would be seen if conducting screening in the primary care clinic. The positive and negative predictive values are influenced by underlying population prevalence and may be lower in a general primary care sample. Additionally, the PHQ-9 was administered via a phone interview which may have resulted in different responses than if it had been self-administered. Finally, the DISC-IV asks questions about a one-month time period while the PHQ-9 asks about the prior 2 weeks. Some of the lack of sensitivity and specificity may be due to these time window differences.

Despite these limitations, the PHQ-9 is a promising screening tool for use among adolescents. It is brief, easy for patients to understand, simple to score, and available without cost. An additional major advantage of the PHQ-9 is that many primary care providers are already using it for the adult population and thus have familiarity with administration and scoring. It performs well in this age group and will be particularly useful for providers or

researchers who want to conduct rapid screening in primary care settings or as part of research protocols.

Acknowledgments

Abbreviations

|	
(PHQ-9)	Patient Health Questionnaire 9-item
(DISC-IV)	Child Diagnostic Interview Schedule
(PHQ-2)	Patient Health Questionnaire 2-item
(DSM-IV)	Diagnostic and Statistical Manual
(GH)	Group Health
(CIS)	Columbia Impairment Scale
(SCARED)	Screen for Child Anxiety Related Emotional Disorders
(PSC-17)	Brief Pediatric Symptom Checklist
(BDI)	Beck Depression Inventory
(MFQ)	Moods and Feelings Questionnaire

Bibliography

1. Screening and treatment for major depressive disorder in children and adolescents: US Preventive Services Task Force Recommendation Statement. Pediatrics. Apr; 2009 123(4):1223–1228. [PubMed: 19336383]

2. Zuckerbrot RA, Maxon L, Pagar D, Davies M, Fisher PW, Shaffer D. Adolescent depression screening in primary care: feasibility and acceptability. Pediatrics. Jan; 2007 119(1):101–108. [PubMed: 17200276]

3. Winter LB, Steer RA, Jones-Hicks L, Beck AT. Screening for major depression disorders in adolescent medical outpatients with the Beck Depression Inventory for Primary Care. J Adolesc Health. 1999; 24(6):389–394. [PubMed: 10401966]

4. Yates P, Kramer T, Garralda E. Depressive symptoms amongst adolescent primary care attenders. Levels and associations. Soc Psychiatry Psychiatr Epidemiol. Jul; 2004 39(7):588–594. [PubMed: 15243698]

5. Gledhill J, Kramer T, Iliffe S, Garralda ME. Training general practitioners in the identification and management of adolescent depression within the consultation: a feasibility study. J Adolesc. Apr; 2003 26(2):245–250. [PubMed: 12581730]

6. Johnson JG, Harris ES, Spitzer RL, Williams JB. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolesc Health. 2002; 30(3):196–204. [PubMed: 11869927]

7. Schubiner H, Tzelepis A, Wright K, Podany E. The clinical utility of the Safe Times Questionnaire. J Adolesc Health. Jul; 1994 15(5):374–382. [PubMed: 7947851]

8. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001; 16(9):606–613. [PubMed: 11556941]

9. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry. Sep-Oct; 2007 29(5):388–395. [PubMed: 17888804]

10. Richardson LP, Rockhill C, Russo JE, et al. Evaluation of the PHQ-2 as a brief screen for detecting major depression among adolescents. Pediatrics. May; 125(5):e1097–1103. [PubMed: 20368315]

11. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIMEMD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. Jama. 1999; 282(18):1737–1744. [PubMed: 10568646]

12. Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME. NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. J Am Acad Child Adolesc Psychiatry. Jan; 2000 39(1):28–38. [PubMed: 10638065]

13. Wells KB, Burnam MA, Leake B, Robins LN. Agreement between face-to-face and telephone-administered versions of the depression section of the NIMH Diagnostic Interview Schedule. J Psychiatr Res. 1988; 22(3):207–220. [PubMed: 3225790]

14. Simon GE, Revicki D, VonKorff M. Telephone assessment of depression severity. J Psychiatr Res. Jul-Sep; 1993 27(3):247–252. [PubMed: 8295157]

15. Bird HR, Andrews H, Schwab-Stone M. Columbia Impairment Scale (CIS). Global measures of impairment for epidemiologic and clinical use with children and adolescents. International Journal of Methods in Psychiatric Research. 1996; 6(4):295–307. al. e.

16. Birmaher B, Brent DA, Chiappetta L, Bridge J, Monga S, Baugher M. Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. J Am Acad Child Adolesc Psychiatry. Oct; 1999 38(10):1230–1236. [PubMed: 10517055]

17. Gardner W, Lucas A, Kolko DJ, Campo JV. Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. J Am Acad Child Adolesc Psychiatry. May; 2007 46(5):611–618. [PubMed: 17450052]

18. Tape, T. Interpreting diagnostic tests. University of Nebraska Medical Center; Available at http://gim.unmc.edu/dxtests/default.htm

19. Angold A, Costello EJ, Messer SC, Pickles A, Winder F, Silver D. The development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. International Journal of Methods in Psychiatric Research. 1995; 5:1–12.

20. Katon W, Russo J, Richardson L, McCauley E, Lozano P. Anxiety and depression screening for youth in a primary care population. Ambul Pediatr. May-Jun; 2008 8(3):182–188. [PubMed: 18501865]

21. Hamilton BE, Minino AM, Martin JA, Kochanek KD, Strobino DM, Guyer B. Annual summary of vital statistics: 2005. Pediatrics. Feb; 2007 119(2):345–360. [PubMed: 17272625]
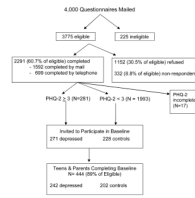
Figure 1.
Recruitment for Adolescent Health Study
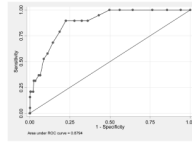
Figure 2.
ROC Curve for PHQ-9 compared to DISC-IV Major Depression Gold Standard

Table 1

Distribution of PHQ-9 Scores According to Depression Diagnostic Status based on the DISC-IV

Level of Depression Severity, PHQ-9 Score	Major Depressive Disorder Last Month (N=19)	Intermediate Depression (N=50)	No Depressive Disorder (N=373)
Minimal, 0-4 (N = 186)	0	6.0%	49.1%
Mild, 5-10 (N = 144)	10.5%	40.0%	32.7%
Moderate, 11-14 (N = 65)	36.8%	38.0%	10.5%
Moderately Severe, 15-19 (N = 34)	31.6%	16.0%	5.4%
Severe, 20-27 (N = 13)	21.1%	0	2.4%

Table 2

Operating Characteristics of Various PHQ-9 Cut-points for diagnosing Major Depression as Determined by the DISC-IV

PHQ-9 Score	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Likelihood Ratio +
≥6	100	50.4	8.3	100	2.01
≥7	94.7	57.0	9.0	99.6	2.20
≥8	89.5	63.8	10.0	99.3	2.47
≥9	89.5	67.6	11.0	99.3	2.76
≥10	89.5	72.1	12.6	99.3	3.21
≥11	89.5	77.5	15.2	99.4	3.98
≥12	79.0	81.3	16.0	98.9	4.23
≥13	68.4	85.6	17.6	98.4	4.75

Area under the ROC Curve: 88% (95%CI = 82% - 94%)

Table 3

Relationship between PHQ-9 Score and Mean Scores for Functional Impairment, Parental Report of Internalizing Symptoms and Total Score on the Pediatric Symptom Check list (N = 442)

Level of Depression Severity, PHQ-9 Score	CIS Score	Symptom Related Difficulty (item 10 on the PHQ-9)	Internalizing Symptoms on the PSC-17 (Parent report)	Total PSC-17 score (Parent report)
0-4	8.4 (6.6)	1.3 (0.4)	2.2 (1.9)	14.4 (9.4)
5-10	16.5 (6.9)	1.8 (0.6)	3.8 (2.2)	20.9 (10.1)
11-14	22.7 (7.4)	2.2 (0.7)	4.8 (2.2)	25.8 (11.0)
15-19	25.0 (7.1)	2.6 (0.8)	5.3 (2.2)	27.4 (11.1)
20-27	27.4 (8.9)	2.6 (0.8)	5.8 (2.7)	27.2 (12.0)
$F_{(4, 437)}$	92.38[***]	19.97[***]	33.23[***]	26.18[***]

[***]
$p < .0001$

[1] Note: Standard deviations are given in parentheses

[2] The recommended cut point clinically significant symptoms on the PSC-17 Internalizing scale is ≥5. The recommended cut point for elevated psychosocial concerns on the total PSC-17 is ≥15.[17]