Evaluation of the PHQ-2 as a Brief Screen for Detecting Major Depression Among Adolescents

Laura P. Richardson, Carol Rockhill, Joan E. Russo, David C. Grossman, Julie Richards, Carolyn McCarty, Elizabeth McCauley and Wayne Katon

*Pediatrics* 2010;125;e1097-e1103; originally published online Apr 5, 2010;
DOI: 10.1542/peds.2009-2712

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://www.pediatrics.org/cgi/content/full/125/5/e1097
Evaluation of the PHQ-2 as a Brief Screen for Detecting Major Depression Among Adolescents

WHAT'S KNOWN ON THIS SUBJECT: The US Preventive Services Task Force recommends screening for depression among adolescents. However, few tools have been validated among adolescents in primary care settings. No studies have examined brief 2-item screening tools among adolescents.

WHAT THIS STUDY ADDS: This study shows that the Patient Health Questionnaire 2-item depression screen has good sensitivity and specificity for detecting major depression among adolescents making it a good candidate for use as a first step in screening for adolescent depression in primary care.

abstract

OBJECTIVE: To examine the validity of the Patient Health Questionnaire 2 (PHQ-2), a 2-item depression-screening scale, among adolescents.

METHODS: After completing a brief depression screen, 499 youth (aged 13–17 years) who were enrolled in an integrated health care system were invited to participate in a full assessment, including a longer depression-screening scale (Patient Health Questionnaire 9-item depression screen) and a structured mental health interview (Diagnostic Interview Schedule for Children). Eighty-nine percent (n = 444) completed the assessment. Criterion validity and construct validity were tested by examining associations between the PHQ-2 and other measures of depression and functional impairment.

RESULTS: A PHQ-2 score of ≥3 had a sensitivity of 74% and specificity of 75% for detecting youth who met Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, criteria for major depression on the Diagnostic Interview Schedule for Children and a sensitivity of 96% and specificity of 82% for detecting youth who met criteria for probable major depression on the Patient Health Questionnaire 9-item depression screen. On receiver operating characteristic analysis, the PHQ-2 had an area under the curve of 0.84 (95% confidence interval: 0.75–0.92), and a cut point of 3 was optimal for maximizing sensitivity without loss of specificity for detecting major depression. Youth with a PHQ-2 score of ≥3 had significantly higher functional-impairment scores and significantly higher scores for parent-reported internalizing problems than youth with scores of <3.

CONCLUSIONS: The PHQ-2 has good sensitivity and specificity for detecting major depression. These properties, coupled with the brief nature of the instrument, make this tool promising as a first step for screening for adolescent depression in primary care. Pediatrics 2010; 125:e1097–e1103

WHAT'S KNOWN ON THIS SUBJECT: The US Preventive Services Task Force recommends screening for depression among adolescents. However, few tools have been validated among adolescents in primary care settings. No studies have examined brief 2-item screening tools among adolescents.

WHAT THIS STUDY ADDS: This study shows that the Patient Health Questionnaire 2-item depression screen has good sensitivity and specificity for detecting major depression among adolescents making it a good candidate for use as a first step in screening for adolescent depression in primary care.
By the age of 18 years, 20% of youth have experienced at least 1 episode of major depression. Depressed youth are at increased risk for suicide, school failure, substance abuse, nicotine dependence, early pregnancy, and social isolation. Younger age of disease onset has been shown to be a predictor of increased risk for delays in mental health treatment, with most adolescents not receiving any treatment until early adulthood. The delay in diagnosis and treatment of mental health disorders in adolescents and an inadequate supply of child mental health specialists has led to increasing focus on screening for depression and improving the quality of depression treatment in pediatric primary care settings.

In response to the growing evidence of effective depression treatments, the US Preventive Services Task Force now recommends screening for depression among adolescents. However, a recent meta-analysis identified only 5 studies with adequate psychometric data for screening of adolescents in primary care. Each of these studies used different instruments, and none examined brief screening questionnaires (ie, 2–3 questions) for depression. Brief screens are important in the primary care setting given the time constraints of busy practices and the need to screen for many health risk behaviors, not just depression.

The Patient Health Questionnaire 2-item depression (PHQ-2) screener is one of the most common brief screens used with adult populations. It has been shown to have good diagnostic validity among multiple large samples of adult primary care patients and comparable sensitivity and specificity to other longer measures of depression. It is often used as a first step in depression screening to identify individuals who require additional evaluation with the remainder of the Patient Health Questionnaire 9-item depression screen (PHQ-9) questions and a clinical interview.

In this study, we evaluated the criteria and construct validity of the PHQ-2 as a screening tool for depressive disorders among adolescents.

**METHODS**

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

**Survey Methods**

Between September 2007 and June 2008, study staff randomly selected 4000 enrollees, aged 13 to 17, who had seen a provider in a GH facility at least once in the previous 12 months. The parents/guardians of selected enrollees received an invitation letter, a consent form, and a brief (10-item) survey for their child. Parents were asked to sign the consent form and give it and the survey to their child to complete in a private place. The child received $2 with the survey. Completion of the survey was taken as a form of assent by the child, and a telephone number for questions was included on all study materials. Parents of youth who did not respond received a second mailing and follow-up telephone calls.

The brief survey consisted of 10 items about age, gender, weight, height, sedentary behaviors, overall health, functional impairment, and depressive symptoms. The PHQ-2 was administered for the first time as a part of this brief questionnaire and was used to determine who would be invited for a follow-up interview. The questions in the PHQ-2 survey asked respondents to rate the frequency they had (1) a depressed mood and/or (2) lack of pleasure in usual activities in the past 2 weeks on a Likert scale of 0 (not at all) to 3 (nearly every day). Scores range from 0 to 6. A score of $\geq 3$ on the PHQ-2 has been found in adults to have the highest sensitivity and specificity and area under the curve in receiver-operating-characteristic (ROC) analysis for a diagnosis of major depression on the basis of a structured psychiatric interview.

A subset of youth ($n = 499$) was invited to participate in the follow-up telephone interview study, during which more in-depth information was obtained on depressive symptoms, functional impairment, and health behaviors. Youth with higher PHQ-2 scores on screening were oversampled such that most youth with a PHQ-2 score of $\geq 3$ ($n = 271$) and a sample of youth with a PHQ-2 score of $\leq 2$ ($n = 228$) who were frequency matched for age and gender were invited to participate. Youth who completed the follow-up interview were mailed $20. Consent for the telephone survey was obtained from both the parent and the child.

**Depression Measures**

The child telephone interview included the PHQ-9 screener and the Diagnostic Interview Schedule for Children depression modules (DISC-IV). The PHQ-2 questions are included as the first 2 questions of the PHQ-9. In our telephone interview, the PHQ-9 was completed before any other depression or mental health measures. We used the PHQ-2 from the telephone interview in
all analyses described in this article to eliminate time between assessments as a reason for disagreement in screening and interview results. The PHQ-9 is a self-administered version of the depression portion of the Primary Care Evaluation of Mental Disorders (PRIME-MD) interview, which uses Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria to assess mental disorders in primary care and was used as a gold standard in 2 of the main primary care–based adolescent screening studies. 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. The DISC-IV is a reliable and valid structured interview designed for lay interviewers and includes algorithms to diagnose DSM-IV disorders in children and adolescents. Telephone versions of structured psychiatric interviews in both adults and youth have been found to have a high correlation with in-person interviews. 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. The 13-item CIS scale measures adolescent impairment in many domains, including school, family, and peer relationships, and has been shown to correlate with the clinician-rated Children’s Global Assessment Scale. To assess anxiety symptoms, youth were asked to complete the brief 5-item version of the Screen for Child Anxiety Related Emotional Disorders (SCARED). With a cutoff of ≥3, the brief SCARED has been shown to have a sensitivity of 74% and a specificity of 73% for discriminating clinically significant anxiety from nonanxiety compared with an interview administered by trained clinicians. To evaluate parent-reported child internalizing symptoms and psychosocial function, 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 on a structured diagnostic interview.

Statistical Analysis
Descriptive statistics were completed for the full sample and stratified according to depression status. Three categories of depression status were used according to algorithms from the DISC-IV: major depression; intermediate depression; or no depressive disorder. Youth were classified as having intermediate depression if they reported at least 3 of the 9 symptom criteria for major depression with or without impairment caused by depression but the diagnostic criteria for major depression were not met. χ² and t test analyses were used to compare categorical and continuous variables, respectively, among depressed and nondepressed individuals on the basis of PHQ-2 score. The individual items in the PHQ-2 as well as the overall PHQ-2 score were examined according to depressive-status category. Subsequently, ROC analyses were performed for the PHQ-2 by using major depression on the DISC-IV as the gold standard.

RESULTS
Of 3775 eligible youth, 2291 (60.7%) completed the brief survey (Fig 1). Twelve percent of the youth (n = 281) screened positive for possible depression with a PHQ-2 ≥3, and 88% screened negative for depressive symptoms. Four hundred ninety-nine youth were invited to participate in the full baseline assessment; 444 (89%) consented, and both the parent and child completed the baseline survey. Two youth who met DSM-IV criteria for bereavement were removed from the analytic sample, resulting in a final sample of 442 youth for the current analysis. Study participants were predominantly female (60%), white (71%), and from urban regions (83%). The mean age of participants was 15.3 years (SD: 1.2 years). The median household income for neighborhoods in which subjects lived was $57,442 (SD: $18,293), and 86% of youth had ≥1 parent who had at least some exposure to higher education. Seven percent of the youth...
were enrolled in a public assistance insurance plan.

Table 1 lists the distribution of scores for each of the individual PHQ-2 items as well as for the full PHQ-2 score according to depression status on DISC-IV (major depression, intermediate depression, or no depression). Youth who met criteria for major depression had significantly higher total scores on the PHQ-2. Youth with “intermediate depression” on the DISC-IV were most likely to report having symptoms on several days but not nearly every day. Youth with no depression diagnosis were most likely to report no symptoms. When the sensitivity and specificity of each individual item in the PHQ-2 was examined, neither individual item performed better than the 2 combined. This result is supported by the significantly greater area under the curve for the PHQ-2 compared with either item (χ² = 39.97; P < .001).

Table 2 lists the test characteristics of the PHQ-2 using the PHQ-9 and DISC-IV as gold standards. The optimal cut point for maximizing sensitivity of the PHQ-2 without loss of specificity was a score of ≥3. At this cut point, the PHQ-2 had a sensitivity of 98.2% for detecting youth with probable major depression according to PHQ-9 criteria and of 73.7% for detecting youth with major depression on the DISC-IV. The specificity was 82.3% for detecting youth with probable major depression on the PHQ-9 and 75.2% for detecting youth with major depression on the DISC-IV. The positive predictive value was 42% for detecting probable major depression on the PHQ-9 and 75.2% for detecting youth with major depression on the DISC-IV. The area under the curve for detecting major depression was 0.84 (95% confidence interval: 0.75–0.92) using the DISC-IV diagnosis as a gold standard and 0.95 (95% confidence interval: 0.93–0.97) using the PHQ-9 diagnosis as a gold standard.

Youth with a PHQ-2 score of ≥3 compared with those with <3 had significantly higher scores for functional impairment as measured by the CIS, as well as parent-reported psychosocial impairment as measured by the PSC-17 (Table 3). In addition, parental reports of internalizing symptoms were significantly higher for those in this group.

Item 9 of the PHQ-9 asks how often the respondent is “thinking that you would be better off dead or that you want to hurt yourself in some way.” Sixteen youth indicated that they had these

**FIGURE 1**
Recruitment for the Adolescent Health Study.

**TABLE 1** Frequency Distribution of Depressed Mood and Anhedonia Items and PHQ-2 Score

<table>
<thead>
<tr>
<th>Item or PHQ-2 Score</th>
<th>Major Depression (n = 19), %</th>
<th>Intermediate Depression (n = 54), %</th>
<th>No Depression (n = 369), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressed mood</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 (not at all)</td>
<td>0</td>
<td>7</td>
<td>53</td>
</tr>
<tr>
<td>1 (several days)</td>
<td>42</td>
<td>56</td>
<td>31</td>
</tr>
<tr>
<td>2 (more than half)</td>
<td>26</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>3 (nearly every day)</td>
<td>32</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Anhedonia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 (not at all)</td>
<td>10</td>
<td>15</td>
<td>56</td>
</tr>
<tr>
<td>1 (several days)</td>
<td>32</td>
<td>54</td>
<td>31</td>
</tr>
<tr>
<td>2 (more than half)</td>
<td>21</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>3 (nearly every day)</td>
<td>37</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>PHQ-2 score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>35</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Totals may not equal 100% because of rounding.
thoughts “more than half the days” or “nearly every day.” Of these 16 youth, 13 (81%) had a PHQ-2 score of ≥3, whereas 3 did not.

The false-positive rate, calculated as 1-specificity, was ~25% when using the DISC-IV as a gold standard or 18% when using the PHQ-9 as a gold standard. To better understand the characteristics of youth with false-positive results, we examined the association between having a false-positive PHQ-2 result and having intermediate depression or a positive screening-test result for another disorder (anxiety or externalizing disorder). Among adolescents with PHQ-2 scores of ≥3 but no DISC-IV diagnosis for major depression (n = 105), 23.3% had intermediate depression on the DISC-IV, 14.7% had major depression in the previous year but not in the previous month, 26.2% had elevated externalizing disorder symptoms (PSC-17 externalizing scale score ≥7), and 55.2% had clinically significant anxiety symptoms (SCARED score ≥3). Taken together, 76.2% of those in the false-positive group had at least 1 of the 4 indications we examined: 39.0% had 1, 32.4% had 2, and 4.8% had 3 of the 4 indications.

**DISCUSSION**

The US Preventive Services Task Force now advises primary clinicians to screen adolescents for depression provided there is a system of care to confirm diagnosis and initiate treatment.\(^16\) To effectively implement broad-based screening in pediatric settings, brief tools are needed. The PHQ-2 is well suited as a first-line screening tool for depression because it is brief, easy to score, and available without cost. When compared with the PHQ-9, the sensitivity and specificity of the PHQ-2 is similar to the Beck Depression Inventory for Primary Care\(^22\) and the adolescent version of the Patient Health Questionnaire.\(^23\) Thus, on the basis of our evaluation, the PHQ-2 has good sensitivity and specificity as a first-line screening tool for adolescents in primary care settings.

Compared with the DISC-IV gold standard, the PHQ-2 has a lower specificity than has been found among adults (75% vs 92%)\(^18\), thus, 25% of youth without disease would have false-positive results compared with only...
Evaluation of the PHQ-2 as a Brief Screen for Detecting Major Depression Among Adolescents

8% of adults screened. In part, this lower specificity may result from the high degree of comorbidity and symptom overlap among youth. The results of our study showed that 76% of the false-positive group had either elevated depressive symptoms that were under the threshold for a major depression diagnosis, had depressive symptoms that had met the cutoff for major depression in the previous year but their current symptoms had improved, or had screening scores suggestive of externalizing or anxiety disorders. Depressive symptoms exist on a continuum, and youth with elevated depressive symptoms are at increased risk for the later development of depression. In addition, youth who have had 1 major depressive episode have a high likelihood of relapse or recurrence. Finally, youth who meet criteria for externalizing and anxiety disorders are at increased risk for the development of major depression and may benefit from monitoring of depressive symptoms over time. Thus, results of this study indicate that youth with false-positive PHQ-2 scores are likely to be at risk for subsequent major depression episodes and might benefit from additional 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. We chose not to change the wording of the PHQ-2 to allow for the use of a single form for settings in which both adolescents and adults are seen. Because we did not include this item, we were not able to determine how it may have modified the performance of the PHQ-2.

On the basis of our findings, we would recommend using a cut point of ≥3 to indicate the need for additional evaluation for depression, particularly if requiring a full diagnostic assessment on all positive youth. However, providers should be aware that a negative screen does not rule out a disorder. The sensitivity of 74% implies that 26% of youth with a depressive disorder would be missed when using this cut point. If providers are concerned about missing depressed youth, they might reasonably choose to use a cut point of ≥2 to maximize sensitivity and then follow it with a longer screening instrument (such as the PHQ-9 or the Beck Depression Inventory) to improve specificity and reduce the number of youth requiring full assessment. If opting for the second approach, it would be wise to also screen for anxiety, given the high rate of elevated anxiety symptoms in youth with false-positive PHQ-2 screening results. In addition, in our study, −20% of youth with suicidal ideation identified by their responses on the PHQ-9 would have been missed by screening with the PHQ-2 alone. Any screening procedures that rely on the PHQ-2 might also include a question about suicidality.

This study has 4 main limitations. First, the study was conducted in an insured population of adolescents in the Pacific Northwest and may not be generalizable to all adolescent populations. Second, the PHQ-2 measure used in this study was a part of the PHQ-9. Thus, one would expect a high degree of correlation of results, and the sensitivity and specificity may be slightly lower if they were administered separately. Although this is a limitation, we feel that our usage parallels common clinical and research practice of administering a screening measure and immediately after the screener with a more definitive measure for those who screen positive. Third, 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. We were very encouraged by the 89% participation rate in the follow-up interview study; however, it is possible that youth who chose not to participate were different from those who did. Finally, questions in the DISC-IV ask about a 1-month period, whereas the PHQ-2 asks about the previous 2 weeks. Some of the lack of sensitivity and specificity may result from these time-window differences.

Despite these limitations, the PHQ-2 is a promising screening tool for use among adolescents. It performs well in this age group and will be particularly useful for providers or researchers who want to conduct a quick initial depression screening in primary care settings or as part of Web-based or survey screening for health risks.

ACKNOWLEDGMENTS

This study was supported by grants from the Group Health Community Foundation Child and Adolescent Grant Program, the University of Washington Royalty Research Fund, a Seattle Children’s Hospital Steering Committee Award, and a K23 award for Dr Richardson from the National Institute of Mental Health (5K23 MH069814-01A1).

REFERENCES


4. Kovacs M. Presentation and course of major depressive disorder during childhood
15. Cheung AH, Zuckerbrodt RA, Jensen PS, Ghaib K, Laraque D, Stein RE. Guidelines for Adolescent Depression in Primary Care (GLAD-PC): II. Treatment and ongoing management. Pediatrics. 2007;120(5). Available at: www.pediatrics.org/cgi/content/full/120/5/e1313