Performance of the PHQ-9 as a Screening Tool for Depression After Stroke

Linda S. Williams, MD; Edward J. Brizendine, MS; Laurie Plue, MA; Tamilyn Bakas, DNS, RN; Wanzhu Tu, PhD; Hugh Hendrie, MD; Kurt Kroenke, MD

Background and Purpose—The purpose of this study was to examine the performance of the Patient Health Questionnaire (PHQ)-9, a 9-item depression scale, as a screening and diagnostic instrument for assessing depression in stroke survivors.

Methods—As part of a randomized treatment trial for poststroke depression (PSD), subjects with and without PSD completed the PHQ-9, a 9-item summed scale, with scores ranging from 0 (no depressive symptoms) to 27 (all symptoms occurring daily). Subjects endorsing 2 or more symptoms of depression were administered the criterion standard Structured Clinical Interview for Depression (SCID). Receiver operating characteristic analysis was used to examine the sensitivity and specificity of the PHQ-9.

Results—Of 316 subjects enrolled, 145 met SCID criteria for major depression or other depressive disorder, and 171 were not depressed. PHQ-9 scores discriminated well between subjects with any versus no depressive disorder, with an area under the curve (AUC) of 0.96, as well as between subjects with and without major depression (AUC=0.96). The AUC was similar regardless of patient age, gender, or ethnicity. A PHQ-9 score ≥10 had 91% sensitivity and 89% specificity for major depression, and 78% sensitivity and 96% specificity for any depression diagnosis.

Conclusions—The PHQ-9 performs well as a brief screener for PSD with operating characteristics similar or superior to other depression measures and similar to its characteristics in a primary care population. Moreover, PHQ-9 scores discriminate equally well between those with and without PSD regardless of age, gender, or ethnicity. (Stroke. 2005;36:635-638.)

Key Words: depression stroke

Poststroke depression (PSD) affects approximately one-third of ischemic stroke survivors, is often undiagnosed and inadequately treated, and is associated with increased morbidity and mortality after stroke. Depression screening after stroke is thus important but can be complicated by cognitive and physical symptoms of stroke that may introduce additional variability in assessment of depressive symptoms and depression diagnosis. Although several established depression screening instruments have been validated in stroke cohorts, these scales can be burdensome for patients to complete, require a trained interviewer to administer, and often are designed only for screening and not as a diagnostic depression tool. The Patient Health Questionnaire 9-item depression scale (PHQ-9) is a 9-item self-administered depression screening and diagnostic tool increasingly used in stroke cohorts. Although these scales can be burdensome for patients to complete, require a trained interviewer to administer, and often are designed only for screening and not as a diagnostic tool. The Patient Health Questionnaire 9-item depression scale (PHQ-9) is a 9-item self-administered depression screening and diagnostic tool increasingly used in stroke cohorts. Although it has not been previously validated in patients with PSD. The purpose of this study was to examine the performance of the PHQ-9 as a screening and diagnostic instrument for assessing depression in ischemic stroke survivors.

Subjects and Methods
Subjects were patients enrolled in the National Institute for Neurologic Disorders and Stroke-funded AIM (Activate, Initiate treatment, Monitor) PSD study. The AIM study consists of a randomized clinical trial of case-management intervention versus usual care in depressed subjects, nested within a longitudinal cohort study that includes nondepressed subjects. Nondepressed subjects were matched 1:1 by site of enrollment to depressed patients. Eligible patients at 4 Indianapolis hospitals were screened for PSD between 1 and 2 months after ischemic stroke. Patients with more than moderate aphasia (National Institutes of Health Stroke Scale language item score >1) or cognitive impairment (modified 6-item Mini-Mental Status score <3) were excluded. We used previously validated methodology to estimate stroke severity at the time of stroke admission from the admission physical examination note. The local human subjects review board approved this study.

All subjects were screened for depression with the PHQ-9, a 9-item scale that assesses the 9 Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) depression symptom criteria for fre-
frequency of occurrence during the previous 2 weeks. The first 2 PHQ-9 items assess the 2 cardinal DSM-IV symptoms of depression: depressed mood and anhedonia. The PHQ-9 can be used as a screening tool, with summed score ranging from 0 (no depressive symptoms) to 27 (all symptoms occurring daily). In this use, a PHQ-9 score ≥10 has been found to have 88% sensitivity and 88% specificity for a diagnosis of major depression. The PHQ-9 can also be used as a diagnostic assessment, with major depression diagnosed if 5 or more of the 9 symptoms have been present at least most of the days of the past 2 weeks and 1 of these symptoms is either depressed mood or anhedonia.

Subjects endorsing any occurrence in the past 2 weeks of at least 2 symptoms on the PHQ-9 or any subject endorsing the depressed mood or anhedonia item were administered the Structured Clinical Interview for Depression (SCID), considered a criterion standard for DSM-IV depressive disorder diagnoses in clinical research. Subjects who did not endorse the mood, anhedonia, or thoughts of death items and who endorsed only 1 other PHQ-9 item were considered nondepressed. SCID diagnoses of major depression and other depression were determined according to standard scoring algorithms in which all depressed subjects must endorse depressed mood and/or anhedonia, and with major depression defined as a total of 5 or more depressive symptoms endorsed and other depression as a total of 3 or 4 depressive symptoms endorsed in the past 2 weeks. Only baseline PHQ-9 and PHQ-2 scores were used in these analyses.

The PHQ-2 is an abbreviated version of the PHQ-9. It is defined as the sum of the anhedonia and mood items of the PHQ-9 and can be used as a very brief depression screening tool in primary care. In primary care patients, a PHQ-2 score ≥3 has 83% sensitivity and 92% specificity for identifying patients with major depression. Because very brief screening tools may be preferred in some clinical settings, we also examined the sensitivity and specificity of the PHQ-2 score for identifying patients with major and any depression, again using the SCID criterion standard for depressive disorder diagnosis.

We used receiver operating characteristic analyses to assess the discriminatory power of the PHQ-9 as a screening and diagnostic tool for any depression (other depression or major depression) and major depression. We also examined the effect of age (60 years or older versus younger than 60 years), gender, and ethnicity (white versus nonwhite) on the PHQ-9 operating characteristics. We assessed the sensitivities and specificities of a PHQ-9 screening score ≥10 and a PHQ-2 score ≥3 for identifying subjects with major depression and any depression. Demographic characteristics were compared between depressed and nondepressed subjects using the Pearson χ² test. SAS version 8.2 (SAS Institute, Cary, NC) was used for all analyses.

Results

Of 316 patients enrolled, 145 were depressed and 171 were nondepressed. There was a greater proportion of younger patients (younger than 60 years old) among the depressed subjects than the nondepressed subjects (50% versus 35% younger than 60 years old; P=0.009) but the groups were similar in terms of gender (57% versus 46% female; P=0.07) and ethnicity (60% versus 58% white; P=0.71). National Institutes of Health Stroke Scale score at the time of stroke was also similar between depressed and nondepressed subjects (mean score 3.3 versus 3.1; P=0.475). All symptoms were endorsed significantly more frequently by depressed than by nondepressed patients (Table 1). Fatigue was the symptom most frequently endorsed by both the depressed (73%) and the nondepressed (32%) subjects. The ninth item of the PHQ-9 (asking whether the patient had been bothered by “thoughts that you would be better off dead or of hurting yourself in some way”) was endorsed by 10% of depressed subjects.

The PHQ-9 had excellent discriminatory power (Figure) for subjects with any depression, with an area under the curve of 0.96, as well as those with major depression only (area under the curve=0.96). The area under the curve was similar between older (0.97) and younger (0.94) subjects, men (0.97) and women (0.94), or white (0.96) and nonwhite (0.96) subjects (Table 2). PHQ-9 scores ≥10 had 91% sensitivity and 89% specificity for major depression and 78% sensitivity and 96% specificity for any depression diagnosis (Table 3). A PHQ-2 score ≥3 had an 83% sensitivity and 84% specificity for major depression and a 78% sensitivity and 95% specificity for any depression diagnosis.

Discussion

The PHQ-9 performed with similarly high diagnostic accuracy for both major depression and any depression in patients with PSD, and performed as well as in stroke survivors as it has in the general medical outpatient population in which it was developed. Performance of the PHQ-9 did not differ by age, ethnicity, or gender. The PHQ-2 also performed quite well as a depression-

### Table 1. Frequency (%) of Patients Endorsing Each PHQ-9 Symptom More Than Half the Days or Every Day

<table>
<thead>
<tr>
<th>Symptom More Than Half the Days or Every Day</th>
<th>Depressed (n=145)</th>
<th>Nondepressed (n=171)</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anhedonia</td>
<td>93 (64)</td>
<td>4 (2)</td>
<td>75 (26, 213)</td>
</tr>
<tr>
<td>Depressed mood</td>
<td>99 (68)</td>
<td>9 (5)</td>
<td>39 (18, 82)</td>
</tr>
<tr>
<td>Trouble sleeping</td>
<td>86 (60)</td>
<td>32 (19)</td>
<td>6 (4, 11)</td>
</tr>
<tr>
<td>Feeling tired</td>
<td>106 (73)</td>
<td>55 (32)</td>
<td>6 (4, 9)</td>
</tr>
<tr>
<td>Change in appetite</td>
<td>64 (44)</td>
<td>20 (12)</td>
<td>6 (3, 11)</td>
</tr>
<tr>
<td>Guilt/worthlessness</td>
<td>67 (46)</td>
<td>8 (5)</td>
<td>18 (8, 38)</td>
</tr>
<tr>
<td>Trouble concentrating</td>
<td>57 (39)</td>
<td>10 (6)</td>
<td>10 (5, 21)</td>
</tr>
<tr>
<td>Feeling slowed down or restless</td>
<td>63 (43)</td>
<td>21 (12)</td>
<td>5 (3, 10)</td>
</tr>
<tr>
<td>Suicidality/thoughts of death</td>
<td>14 (10)</td>
<td>0 (0.0)</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA indicates not applicable.
screwing tool with nearly identical performance to the PHQ-9 in identifying subjects with any depression. However, for diagnosis and more complete clinical evaluation of depression symptoms, those scoring ≥3 on the PHQ-2 should be administered the additional 7 items to complete the PHQ-9.

Like previous studies,19,20 we also found much higher endorsement of all depression symptoms in patients with depression compared with those without. Even for the more somatic symptoms of depression (agitation, sleep disturbance, fatigue, and appetite disturbance), depressed patients were more than twice as likely to endorse these symptoms compared with those without depression. This finding speaks to the necessity to actively screen for depression in the poststroke period rather than attributing physical symptoms to the stroke itself. Importantly, although 10% of our sample endorsed the “suicidality/thoughts of death” item, a proportion identical to that reported in an earlier study of suicidal thoughts at 3 and 12 months after stroke,21 our standardized suicidality assessment demonstrated that endorsement of this item almost exclusively represented passive thoughts (ie, increased mortality or likelihood of death, or thoughts of whether life was worth living) rather than active suicidal ideation.

Other depression screening tools have been validated in stroke populations6–10 and at least 1 scale has been developed for depression screening in patients with aphasia.22 Although most of these scales have performed reasonably well in stroke cohorts, they are longer than the PHQ-9, and many can only be used for depression screening, not for specific depressive disorder diagnoses. A key difference between the PHQ-9 and other depression scales is that it is based on the 9 DSM-IV symptoms of depression, which facilitates its use as a depression diagnostic tool as well as a screening instrument. Further, although other scales may perform differently in different ethnic groups or among patients of different age or gender,23,24 this critical aspect of scale measurement is infrequently assessed. We found that the PHQ-9 performed equally well regardless of age, gender, or ethnicity, suggesting that it can be confidently used even in a heterogeneous group of patients. Finally, the PHQ-9 has also been demonstrated to be sensitive to change in assessing depression outcomes over time and thus is valuable for monitoring response to depression therapy,25,26 a key aspect of scale performance that we plan to evaluate in this study cohort.

A potential limitation of our study is the inherent differences between clinical trial cohorts and the entire population with a given condition. This study included patients (both depressed and nondepressed) who were participants in a clinical trial and so are likely to have different physical, psychological, and behavioral characteristics than a population-based stroke cohort. Further, these were patients with no more than moderate language or cognitive effects of stroke and most of whom were outpatients at the time of evaluation. Although it is encouraging that we observed no differences in PHQ-9 performance by age, gender, or ethnicity, it would be beneficial to evaluate the PHQ-9 in other stroke samples to ensure its measurement characteristics are stable across the full range of stroke populations and severity levels. Additionally, in some subjects, the same rater administered both the PHQ-9 and the SCID, which may contribute to the observed agreement between the scales. We addressed this possibility by conducting rigorous training and evaluation of all study personnel to minimize interviewer bias, but we acknowledge that for the purposes of comparison these 2 scales would have always been scored by independent interviewers.

Because PSD is common and is associated with increased morbidity and mortality after stroke,27–29 systematic screening for depression symptoms should be considered in all stroke survivors in the first months after stroke. Symptoms of PSD may be misattributed by the patient or the health care provider as expected physical effects of stroke, or they may be missed in the context of a busy follow-up visit in which review of diagnostic studies and attention to secondary stroke prevention often take precedence. As with all screening tools, a few false-positive and false-negative assignments are expected; if the clinician feels the patient might still be depressed despite not achieving the threshold score on the PHQ-9, additional questions should be asked to clarify the diagnosis. However, the use of a brief but accurate depression screening tool like the PHQ-9 that patients can self-complete before the provider visit could help increase

### TABLE 2. Receiver Operating Curve Analyses by Age, Gender, and Ethnicity

|                | AUC
|----------------|---
| Any depression | 0.96
| Major depression | 0.96
| Age            | 0.96
| Older than 60 y | 0.97
| 60 y or younger | 0.94
| Gender         | 0.97
| Men            | 0.97
| Women          | 0.94
| Ethnicity      | 0.96
| White          | 0.96
| Nonwhite       | 0.95

AUC indicates area under the curve, with scores ranging from 0.50 (indicating chance agreement) to 1.0 (perfect agreement).

### TABLE 3. Sensitivity and Specificity of the PHQ-9 and PHQ-2 as a Screening and Diagnostic Tool

<table>
<thead>
<tr>
<th></th>
<th>PHQ-9 ≥10</th>
<th>PHQ-2 ≥3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major depression</td>
<td>90.6</td>
<td>83.0</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>96/106</td>
<td>88/106</td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(85.0, 96.1)</td>
<td>(75.9, 90.2)</td>
</tr>
<tr>
<td>Specificity</td>
<td>88.6</td>
<td>83.8</td>
</tr>
<tr>
<td>Test negative/SCID negative</td>
<td>186/210</td>
<td>176/210</td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(84.3, 92.9)</td>
<td>(78.8, 88.8)</td>
</tr>
<tr>
<td>Any depression</td>
<td>77.9</td>
<td>77.9</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>113/145</td>
<td>113/145</td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(71.2, 84.7)</td>
<td>(71.2, 84.7)</td>
</tr>
<tr>
<td>Specificity</td>
<td>95.9</td>
<td>94.7</td>
</tr>
<tr>
<td>Test negative/SCID negative</td>
<td>164/171</td>
<td>162/171</td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(92.9, 98.9)</td>
<td>(91.4, 98.1)</td>
</tr>
</tbody>
</table>
### TABLE 4. PHQ-9 Depression Severity Scale

<table>
<thead>
<tr>
<th>Over the last 2 weeks, how often have you been bothered by any of the following problems?</th>
<th>Not at All</th>
<th>Several Days</th>
<th>More Than Half the Days</th>
<th>Nearly Every Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Little interest or pleasure in doing things</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2. Feeling down, depressed, or hopeless</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3. Trouble falling or staying asleep, or sleeping too much</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4. Feeling tired or having little energy</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5. Poor appetite or overeating</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6. Feeling bad about yourself—or that you are a failure or have let yourself or your family down</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7. Trouble concentrating on things, such as reading the newspaper or watching television</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8. Moving or speaking so slowly that other people could have noticed, or the opposite—being so fidgety or restless that you have been moving around a lot more than usual</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9. Thoughts that you would be better off dead or of hurting yourself in some way</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

From the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MD PHQ). The PHQ was developed by Drs Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke, and colleagues. For research information, contact Dr Spitzer at rls@nysicp.org.

PHQ-9 Copyright © 1999 Pfizer Inc. All rights reserved. Reproduced with permission. PRIME MD TODAY is a trademark of Pfizer Inc.

## References


Performance of the PHQ-9 as a Screening Tool for Depression After Stroke
Linda S. Williams, Edward J. Brizendine, Laurie Plue, Tamilyn Bakas, Wanzhu Tu, Hugh Hendrie and Kurt Kroenke

Stroke. 2005;36:635-638; originally published online January 27, 2005;
doi: 10.1161/01.STR.0000155688.18207.33
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2005 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/36/3/635

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/