Research report

The PHQ-8 as a measure of current depression in the general population[☆]

Kurt Kroenke [a,*], Tara W. Strine [b], Robert L. Spitzer [c], Janet B.W. Williams [c], Joyce T. Berry [d], Ali H. Mokdad [b]

[a] *Department of Medicine, Indiana University School of Medicine and Regenstrief Institute, Indianapolis, IN, United States*
[b] *Centers for Disease Control and Prevention, Atlanta, GA, United States*
[c] *Department of Psychiatry, Columbia University, and New York State Psychiatric Institute, New York, NY, United States*
[d] *Substance Abuse and Mental Health Services Administration, DC, United States*

Abstract

Background: The eight-item Patient Health Questionnaire depression scale (PHQ-8) is established as a valid diagnostic and severity measure for depressive disorders in large clinical studies. Our objectives were to assess the PHQ-8 as a depression measure in a large, epidemiological population-based study, and to determine the comparability of depression as defined by the PHQ-8 diagnostic algorithm vs. a PHQ-8 cutpoint ≥ 10.
Methods: Random-digit-dialed telephone survey of 198,678 participants in the 2006 Behavioral Risk Factor Surveillance Survey (BRFSS), a population-based survey in the United States. Current depression as defined by either the DSM-IV based diagnostic algorithm (i.e., major depressive or other depressive disorder) of the PHQ-8 or a PHQ-8 score ≥ 10; respondent sociodemographic characteristics; number of days of impairment in the past 30 days in multiple domains of health-related quality of life (HRQoL).
Results: The prevalence of current depression was similar whether defined by the diagnostic algorithm or a PHQ-8 score ≥ 10 (9.1% vs. 8.6%). Depressed patients had substantially more days of impairment across multiple domains of HRQoL, and the impairment was nearly identical in depressed groups defined by either method. Of the 17,040 respondents with a PHQ-8 score ≥ 10, major depressive disorder was present in 49.7%, other depressive disorder in 23.9%, depressed mood or anhedonia in another 22.8%, and no evidence of depressive disorder or depressive symptoms in only 3.5%.
Limitations: The PHQ-8 diagnostic algorithm rather than an independent structured psychiatric interview was used as the criterion standard.
Conclusions: The PHQ-8 is a useful depression measure for population-based studies, and either its diagnostic algorithm or a cutpoint ≥ 10 can be used for defining current depression.
© 2008 Elsevier B.V. All rights reserved.

Keywords: Depression; Psychometrics; Prevalence; Epidemiology; Quality of life; Patient Health Questionnaire

---

1. Introduction

Depression is not only the most common mental disorder in general practice as well as mental health settings, but also is a major public health problem. The World Health Organization now recognizes depression as one of the most burdensome diseases in the world (World Health Organization, 2002). It is also among the leading causes of decreased work productivity (Stewart et al., 2003). The prevalence and impact of depression in the United States has been assessed in important population-based studies, with modern methods first used in the Epidemiological Catchment Area study in the early 1980s (Robins and Regier, 1991) and proceeding to the National Comorbidity Survey in 1990–1992 (Kessler et al., 1994) and its replication (NCS-R) a decade later (Kessler et al., 2003). Utilizing structured psychiatric interviews, these landmark epidemiological studies have provided invaluable information on the community prevalence of depression and other mental disorders.

However, there are a number of periodic population-based surveys conducted by federal or state agencies that provide an opportunity for more regular surveillance, although these surveys do not focus exclusively on depression or psychiatric conditions. Because mental health may be only one of a number of health indicators assessed, brief measures may be essential to reduce respondent burden. One increasingly popular measure for assessing depression is the Patient Health Questionnaire nine-item depression scale (PHQ-9). Since its original validation study in 2001 (Kroenke et al., 2001), the PHQ-9 already has been used in several hundred published studies and translated into more than 30 languages. It consists of the nine criteria for depression from the *Diagnostic and Statistical Manual of Mental Disorders, fourth edition* (DSM-IV). The PHQ-9 is half the length of many depression measures, comparable or superior in operating characteristics, and valid as both a diagnostic and severity measure (Lowe et al., 2004a; Williams et al., 2002a,b). It has been used in clinical (Diez-Quevedo et al., 2001; Kroenke and Spitzer, 2002) and population-based settings (Martin et al., 2006) and is valid in self-administered (Diez-Quevedo et al., 2001; Kroenke et al., 2001) and telephone-administered modes (Pinto-Meza et al., 2005). Additionally, the PHQ-9 is effective for detecting depressive symptoms in various racial/ethnic groups (Huang et al., 2006a,b) and older populations (Klapow et al., 2002), as well as in patients with neurological disorders (Bombardier et al., 2006; Bombardier et al., 2004; Callahan et al., 2006; Fann et al., 2005; Williams et al., 2004; Williams et al., 2005), cardiovascular disease (Holzapfel et al., 2007; Ruo et al., 2003), HIV/AIDS

(Justice et al., 2004), diabetes (Glasgow et al., 2004; Katon et al., 2004), chronic kidney disease (Drayer et al., 2006), cancer (Dwight-Johnson et al., 2005), rheumatological disorders (Lowe et al., 2004c; Rosemann et al., 2007), gastrointestinal disease (Persoons et al., 2001), dermatological disorders (Picardi et al., 2004), and other conditions (Lowe et al., 2004b; Maizels et al., 2006; Persoons et al., 2003; Scholle et al., 2003; Spitzer et al., 2000; Tietjen et al., 2007; Turner and Dworkin, 2004; Turvey et al., 2007).

In order to assess the current prevalence and impact of depression in the United States, an eight-item version of the Patient Health Questionnaire depression scale (PHQ-8) recently was made available for use by state health departments in the 2006 Behavioral Risk Factor Surveillance Survey (BRFSS). The PHQ-8 is comparable to the PHQ-9 in terms of diagnosing depressive disorders when using a DSM-IV based diagnostic algorithm (Corson et al., 2004; Kroenke and Spitzer, 2002). However, there is evidence that a PHQ-8 score ≥ 10 represents clinically significant depression (Kroenke et al., 2001) and is more convenient to use than a diagnostic algorithm. In this paper, we compare the standard diagnostic algorithm and the PHQ-8 cutpoint of 10 in terms of depression prevalence, respondent sociodemographic characteristics, PHQ-8 operating characteristics, and construct validity as assessed by multiple domains of health-related quality of life. Assessment of the PHQ-8 in this large, epidemiological study may provide further evidence of its utility as a depression measure in population-based research.

2. Methods

2.1. Behavioral Risk Factor Surveillance Survey (BRFSS)

The BRFSS is a surveillance system operated by state health departments in collaboration with CDC. It aims to collect uniform, state-specific data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and preventable infectious diseases in the adult population (Centers for Disease Control and Prevention, 2005; Mokdad et al., 2003). Trained interviewers collect data from a standardized questionnaire using an independent probability sample of households with telephones in the non-institutionalized U.S. adult population. Data from all states and areas were pooled to produce national estimates.

The BRFSS questionnaire consists of three parts: 1) core questions asked in all 50 states, the District of Columbia (D.C.), Puerto Rico (PR), and the U.S. Virgin Islands (USVI); 2) supplemental modules, which are

series of questions on specific topics (e.g., adult asthma history, intimate partner violence, mental health); and 3) state-added questions. In 2006, trained interviewers administered questions about depression severity and lifetime diagnosis of anxiety and depression (Anxiety and Depression Module) in 38 states as well as D.C., PR, and USVI. Additional BRFSS methodology is described elsewhere (Holtzman, 2004). All BRFSS questionnaires, data, and reports are available at http://www.cdc.gov/brfss.

2.2. Patient Health Questionnaire eight-item depression scale (PHQ-8)

To assess the prevalence of depression and its severity in the general U.S. population, the standardized and validated PHQ-8 (see Appendix A) was used (Kroenke and Spitzer, 2002). The PHQ-8 consists of eight of the nine criteria on which the DSM-IV diagnosis of depressive disorders is based (American Psychiatric Association, 1994). The ninth question in the DSM-IV assesses suicidal or self-injurious thoughts. It was omitted because interviewers are not able to provide adequate intervention by telephone. Research indicates that the deletion of this question has only a minor effect on scoring because thoughts of self-harm are fairly uncommon in the general population, and the ninth item is by far the least frequently endorsed item on the PHQ-9 (Huang et al., 2006a; Kroenke and Spitzer, 2002; Lee et al., 2007; Rief et al., 2004). Indeed, the two original validation studies of the PHQ totaling 6000 patients established that identical scoring thresholds for depression severity could be used for the PHQ-9 and PHQ-8 (Kroenke and Spitzer, 2002).

The PHQ-8 response set was standardized to make it similar to other BRFSS questions by asking the number of days in the past 2 weeks the respondent had experienced a particular depressive symptom. The modified response set was converted back to the original response set: 0 to 1 day="not at all," 2 to 6 days="several days," 7 to 11 days="more than half the days," and 12 to 14 days="nearly every day," with points (0 to 3) assigned to each category, respectively. The scores for each item are summed to produce a total score between 0 and 24 points. A total score of 0 to 4 represents no significant depressive symptoms. A total score of 5 to 9 represents mild depressive symptoms; 10 to 14, moderate; 15 to 19, moderately severe; and 20 to 24, severe. (Kroenke et al., 2001) Current depression was defined in two ways: 1) a PHQ-8 algorithm diagnosis of major depression (this requires either the first or second item (depressed mood or anhedonia) to be present "more than half the days" and at least 5 of the 8 symptoms to be present "more than half the

days") or other depression (2 to 4 symptoms, including depressed mood or anhedonia, are required to be present "more than half the days"); 2) a PHQ-8 score of ≥ 10, which has an 88% sensitivity and 88% specificity for major depression (Kroenke and Spitzer, 2002) and, regardless of diagnostic status, typically represents clinically significant depression (Corson et al., 2004; Kroenke et al., 2001).

2.3. Health-related quality of life and other items

Three health-related quality of life (HRQoL) questions with demonstrated validity and reliability for population health surveillance were examined (Andresen et al., 2003; Mielenz et al., 2006; Moriarty et al., 2003). The three questions involved respondents' self-assessment of their health over the previous 30 days. 1) Physical health: "How many days was your physical health, which includes physical illness or injury, not good?" 2) Mental health: "How many days was your mental health, which includes stress, depression, and problems with emotions, not good?" 3) Activity limitations: "How many days did poor physical or mental health keep you from doing your usual activities, such as self-care, work, or recreation?"

Additionally, a "Healthy Days Symptoms Module" was used in Delaware, Hawaii, and Rhode Island. Questions in this module also referred to the previous 30 days: 1) Depressive symptoms: "How many days did you feel sad, blue, or depressed?" 2) Anxiety symptoms: "How many days did you feel worried, tense, or anxious?" 3) Sleep problems: "How many days have you felt you did not get enough rest or sleep?" 4) Pain limitations: "How many days did pain make it difficult to do your usual activities?" 5) Vitality: "How many days have you felt very healthy and full of energy?" We calculated fatigue by subtracting the number of days of vitality from 30.

Sociodemographic information was obtained for each respondent. Employment status was assessed by the question: "Are you currently: employed for wages, self-employed, out of work for more than 1 year, out of work for less than 1 year, a homemaker, a student, retired, or unable to work?" Additionally, two questions were asked about lifetime diagnosis: "Has a doctor or other health care provider ever told you that you have an anxiety disorder (including acute stress disorder, anxiety, generalized anxiety disorder, obsessive-compulsive disorder, panic attacks, panic disorder, phobia, posttraumatic stress disorder, or social anxiety disorder)?" and "Has a doctor or other health care provider ever told you that you have a depressive disorder (including depression, major depression, dysthymia, or minor depression)?"

There were 198,678 respondents from the 38 states, D.C., PR, and USVI who completed all PHQ-8

questions. Of these, nearly all (198,574, or 99.95%) completed at least one of the first three HRQoL items (196,673 for mental health, 196,141 for physical health, and 197,543 for activity limitations). Among the 13,622 respondents in Delaware, Hawaii, and Rhode Island, 13,619 (99.98%) answered at least one of the 5 HRQOL questions (depression 13,514, anxiety 13,487, sleep 13,536, fatigue 13,381, and pain 13,534). The median cooperation rate of BRFSS (i.e., the percentage of eligible respondents who completed the survey) was 74.5%.

2.4. Analysis

Depression was classified as either major depression or other depression (using the PHQ-8 diagnostic algorithm as described) or a PHQ-8 score ≥ 10. Sociodemographic characteristics of depressed and nondepressed respondents were compared. The frequency distribution of major depression and other depression by standard PHQ-8 severity intervals (0–4, 5–9, 10–14, 15–19, and 20–24) as well as the commonly used cutpoint of ≥ 10 were described (Kroenke and Spitzer, 2002). Operating characteristics (sensitivity, specificity, likelihood ratios) for PHQ-8 intervals and cutpoint were calculated, using the PHQ-8 diagnostic algorithm as the criterion standard (Kroenke and Spitzer, 2002; Kroenke et al., 2001). The mean number of impairment days in the past 30 days for HRQoL domains was determined for depressed and nondepressed groups. Because of the large sample size, statistical testing was not emphasized.

Weighting in BRFSS is designed to make the total number of cases equal to the number of people in the state who are age 18 and older. In the BRFSS, such poststratification serves as an adjustment for noncoverage and nonresponse and forces the total number of cases to equal population estimates for each geographic region, which for the BRFSS is usually a state. Sample characteristics (Table 1) and PHQ-8 operating characteristics (Table 2) use unweighted BRFSS data, while construct validity analyses (Table 3 and Figs. 1 and 2) use weighted data.

3. Results

3.1. Respondent characteristics

Data were analyzed from 198,678 respondents to the 2006 BRFSS survey. Overall, the sample was 61.6% women, 78% non-Hispanic white, 58.3% currently employed, 61.2% college educated, and 56.9% currently married. A lifetime diagnosis of a depressive or anxiety disorder was reported by 18.0% and 12.3%, respectively.

Table 1 compares the characteristics of depressed vs. nondepressed respondents, with depression defined either by the PHQ-8 diagnostic algorithm (major depressive or other depressive disorder) or by a PHQ-8 cutpoint ≥ 10. Two findings should be emphasized. First, depressed respondents were more likely to be women, nonwhite, less educated, unemployed or unable to work, unmarried, and younger than 55 years. Not surprisingly, depressed respondents also were much more likely to report lifetime diagnoses of both depression and anxiety. Second, characteristics of the depressed groups were quite similar between the two methods of defining depression, as were characteristics of the nondepressed groups. Compared to the diagnostic algorithm, the cutpoint method produced slightly lower estimates of depression in men and in the two oldest age groups and modestly higher estimates in those with a self-reported lifetime diagnosis of depression or anxiety.

3.2. PHQ-8 distribution and operating characteristics

Table 2 shows the relationship between PHQ-8 severity scores and depression diagnostic status. There were 8476 respondents with major depression in the BRFSS sample using the PHQ-8 diagnostic algorithm, resulting in a prevalence of 4.3%. There were 18,053 respondents with any depression using the diagnostic algorithm and 17,040 with any depression using a PHQ-8 cutpoint of ≥ 10, yielding relatively similar prevalences of 9.1% and 8.6%, respectively. No respondents with scores less than 10 had major depression, because this diagnosis requires at least 5 symptoms to be present more than half the days (resulting in a score of 2 for each symptom).

The sensitivity of a PHQ-8 score ≥ 10 is the proportion of respondents with a depressive disorder who have a score of 10 or greater, and the specificity is the proportion of respondents without a depressive disorder who have a score less than 10. The sensitivity and specificity of a PHQ-8 score ≥ 10 for major depressive disorder (vs. other + none) were 100% (8476/8476) and 95% (181,638)/(190,202), respectively; for any depressive disorder, the sensitivity and specificity were 70% (12,556/18,053) and 98% (176,141/180,625). We also calculated the likelihood ratios associated with the PHQ-8 score ranges or thresholds shown in Table 3. The likelihood ratio is defined as the ratio of the probability of a score range or threshold count in individuals with and without a depressive disorder. For example, 9968 people had PHQ-8 scores of 10–14. The likelihood ratio associated

Table 1
Characteristics of respondents by depression status — percent in various groups

Characteristic	No depressive disorder (n=180,625)	PHQ-8<10 (n=181,638)	Depressive disorder [a] (n=18,053)	PHQ-8 ≥10 (n=17,040)
Sex				
Women	61.0	60.7	67.4	71.3
Men	39.0	39.3	32.6	28.7
Age				
18–24	4.5	4.5	5.9	6.2
25–34	12.4	12.4	12.3	13.1
35–44	17.6	17.5	18.3	19.6
45–54	21.1	21.0	24.7	26.4
55–64	19.9	19.8	20.2	20.5
65–74	14.0	14.2	10.6	8.6
75 or greater	10.5	10.7	8.0	5.6
Race/ethnicity				
White	78.8	78.6	69.6	70.9
Black	7.7	7.8	11.4	10.7
Hispanic	7.6	7.7	11.1	10.2
Other	5.9	5.9	7.9	8.1
Education				
<High school	8.5	8.6	19.4	18.7
High school	28.6	28.7	36.6	35.5
≥College	62.9	62.7	44.0	45.8
Employment status				
Employed	60.0	60.0	41.0	40.4
Retired	23.0	23.3	16.4	13.1
Homemaker or student	10.0	10.0	10.0	10.3
Unable to work	3.7	3.6	23.7	26.6
Unemployed	3.2	3.2	9.0	9.6
Marital status				
Currently married	58.6	58.5	40.9	40.3
Previously married	26.6	26.7	39.5	39.4
Never married	14.9	14.8	19.6	20.3
Lifetime diagnoses				
Depression	14.7	14.1	51.0	59.7
Anxiety	9.9	9.4	36.8	43.3

[a] Depressive disorder is *major depression* or *other depression* according to PHQ-8 algorithm.

with a score in this range was $(5922/18,053) \div (4046/180,625) = 15$. This means that a PHQ-8 score of 10–14 is 15 times more likely to occur in persons with a depressive disorder than in persons without a depressive disorder. The likelihood ratios for any depressive disorder for PHQ-8 scores of 0–4, 5–9, 10–14, 15–19, and 20–24 were 0.01, 2, 15, 103, and 2237. The likelihood ratio for a PHQ-8 score ≥10 was 28.

Some of the DSM-IV criteria for depressive disorders are nonspecific, particularly the somatic symptoms (fatigue, trouble sleeping, and appetite or weight changes). They are frequent physical complaints in the general population, arising not only from depression, but also from a number of medical conditions. They also exist as idiopathic symptoms. These symptoms could contribute to elevated PHQ-8 scores and account for false–positive cases of depression when using a PHQ-8

cut point (≥10) rather than the diagnostic algorithm. On the other hand, two DSM-IV criteria – depressed mood and anhedonia – are specific to depressive disorders: At least one of these two symptoms is required for diagnosis. Of the 4484 persons with a PHQ-8 score ≥10 who did not meet DSM-IV criteria for either a major or other depressive disorder, 597 (13.3%) scored zero on both depressed mood and anhedonia. The rest scored a 1 or greater on either or both symptoms, meaning that they experienced at least 2–6 days of one or both core depressive symptoms in the past 2 weeks. Of the 17,040 respondents with a PHQ-8 score ≥10, major depressive disorder was present in 8476 (49.7%), other depressive disorder in 4080 (23.9%), depressed mood or anhedonia in 3887 (22.8%), and no evidence of depressive disorder or depressive symptoms in 597 (3.5%).

Table 2
Distribution of depressive symptom severity and depressive disorders in BRFSS

Depressive symptom severity	Depressive disorder by diagnostic algorithm				Total
[PHQ-8 Score]	Major	Other	Any	None	
	(n=8476)	(n=9577)	(n=18,053)	(n=180,625)	(n=198,678)
0–4	0	200	200	149,921	150,121
5–9	0	5297	5297	26,220	31,517
10–14	1944	3978	5922	4046	9968
15–19	4296	102	4398	428	4826
20–24	2236	0	2236	10	2246
≥10	8476	4080	12,556	4484	17,040
<10	0	5497	5497	176,141	181,638

3.3. HRQoL impairment and depression

Table 3 summarizes the number of days in the past 30 days that respondents reported impairment in various domains of HRQoL. There are several key findings. First, compared to nondepressed respondents, those who are depressed reported substantially more days of impairment in all domains of HRQoL. Not surprisingly, depressed patients experienced the greatest impairment in depressive and anxiety symptoms, somatic symptoms that are core criteria for depression (fatigue and sleep problems), and mental health, reporting 10–20 days of impairment in each of these domains in the past 30 days. They also reported limitations during 10 of the past 30 days due to physical problems or pain.

A second key finding is that the number of days of impairment in both depressed and nondepressed individuals does not differ by definition (PHQ-8 diagnostic algorithm vs. a PHQ-8 cutpoint of 10). For every domain, impairment is equal or slightly greater for depression defined by a PHQ-8 score ≥ 10 (Figs. 1 and 2).

4. Discussion

BRFSS provided an excellent opportunity to examine the PHQ-8 in a large, representative study of the U.S. population. The two methods of estimating current depression – the validated PHQ-8 diagnostic algorithm based upon DSM-IV criteria, and a PHQ-8 cutpoint of 10 – yielded similar prevalences (9.1% and 8.6%, respectively). Also, sociodemographic characteristics were similar in the depressed and nondepressed groups defined by these two methods. Patients classified by the PHQ-8 as having current depression had substantial impairment across multiple domains of HRQoL, and impairment was nearly identical whether groups were classified by the diagnostic algorithm or cutpoint approach. Finally, 96.5% of respondents with a PHQ-8 score ≥ 10 had either a depressive disorder or core depressive symptoms.

The population prevalence of current depression detected by the PHQ-8 in BRFSS is similar to those reported in some other population-based studies, and slightly lower than those reported in other studies. The

Table 3
Relationship between impairment in health-related quality of life and depression status as determined by PHQ-8 diagnostic algorithm and cutpoint

Health-related quality of life domain	Mean number (SEM) of impaired days in past 30 days in specific health-related quality of life domains			
	No depressive disorder [a]	PHQ-8 <10	Depressive disorder [a]	PHQ-8 ≥10
Physical health not good	2.78 (0.04)	2.75 (0.04)	10.09 (0.19)	10.76 (0.20)
Mental health not good	2.27 (0.03)	2.23 (0.03)	14.37 (0.20)	15.34 (0.21)
Activity limited days	1.39 (0.03)	1.36 (0.03)	9.49 (0.18)	10.23 (0.19)
Depressive symptoms [b]	1.44 (0.05)	1.43 (0.04)	13.70 (0.52)	15.57 (0.51)
Anxiety symptoms [b]	3.24 (0.08)	3.21 (0.09)	15.15 (0.55)	17.25 (0.55)
Fatigue [b]	9.33 (0.12)	9.30 (0.12)	20.98 (0.50)	23.13 (0.36)
Sleep problems [b]	6.98 (0.12)	6.93 (0.12)	17.04 (0.52)	19.17 (0.51)
Pain-limited days [b]	2.14 (0.07)	2.14 (0.07)	9.16 (0.46)	10.19 (0.51)

[a] Depressive disorder is *major depression* or *other depression* according to PHQ-8 algorithm.
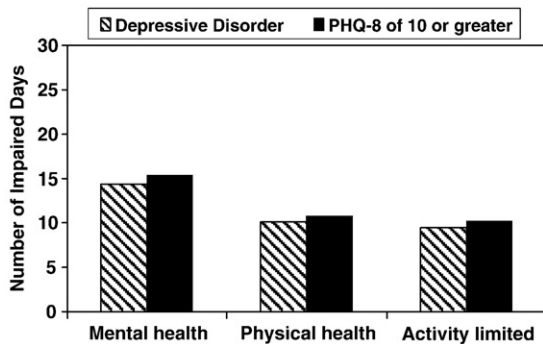[b] Questions asked only in Delaware, Hawaii, and Rhode Island.

Fig. 1. Number of self-reported impairment days out of the past 30 days in the domains of mental health, physical health, and activity limitations. Data available from 198,678 BRFSS respondents from 38 states, the District of Columbia, Puerto Rico, and U.S. Virgin Islands.

National Comorbidity Survey in the early 1990s, and its replication a decade later, revealed that the 12-month prevalence of a major depressive episode in the U.S. population ranges from 6.6% to 10.3% (Kessler et al., 1994; Kessler et., 2003). A population-based study in Germany using the PHQ-9 reported major depression in 3.8% and other depression in 5.4% of respondents (vs. 4.3% and 4.8% in the BRFSS sample) (Martin et al., 2006). The same study reported that 7.2% of respondents had a PHQ-9 score ≥ 10, vs. 8.6% of BRFSS respondents (Rief et al., 2004). A population-based study in Australia using the PRIME-MD (an earlier interviewer-administered version of the PHQ) found a 7% prevalence of major depression and an 18% prevalence of any depression (Hawthorne et al., 2003). Our BRFSS findings of current depression by the PHQ-8 diagnostic algorithm (9.1%) or cutpoint of 10 (8.6%) suggest that rates using the PHQ-8 probably represent accurate or slightly conservative estimates of depression prevalence in the United States.

There are several large studies confirming the validity of the PHQ-8 as both a diagnostic and severity measure in clinical populations (Corson et al., 2004; Kroenke and Spitzer, 2002). Our study is the first to test the PHQ-8 in a large population-based survey. The PHQ-8 may be useful in some types of research, particularly epidemiological/population-based studies, postal or web-based surveys, and clinical studies in which depression is a secondary outcome. In such instances, depression prevalence and severity are expected to be low, the 9th item of the PHQ-9 is infrequently endorsed and even then usually represents passive thoughts of death rather than suicidal ideation, and immediate mental health back-up to interview all the false–positive endorsements of the 9th item is not feasible. In contrast, the PHQ-9 would be preferred in most settings where depression is being diagnosed and

treated because of the importance of recognizing those uncommon but serious instances of suicidal ideation. Also, a modeling study of general practitioners' decision-making suggested that potential suicidal ideation was one of the two most important cues (the other being duration of depressive symptoms) for prescribing antidepressants (Smith and Gilhooly, 2006).

The comparison in BRFSS of two different methods of scoring the PHQ-8 also has broader implications regarding the current debate about a categorical vs. dimensional approach to psychiatric diagnoses. In fact, it is likely that revisions in DSM-V (and possibly ICD 11) will emphasize the complementary value of adding dimensional assessment to the categorical diagnoses of depression as well as other psychiatric disorders (Helzer et al., 2006). Likewise, Kendall and Jablensky argue that while discrete boundaries within and between psychiatric diagnoses have several limitations in terms of validity, coupling dimensional assessment with categorical diagnoses may have utility in monitoring response to treatment, predicting outcomes (e.g., recovery, remission, relapse), comparing different populations, evaluating the impact of social and occupational functioning, and investigating potential etiology (Kendall and Jablensky, 2003). The categorical and dimensional scoring of the PHQ-9 and PHQ-8 may therefore be advantageous.

Systematic reviews have shown that the PHQ-9 has operating characteristics comparable to other depression measures, including the Hospital Anxiety and Depression (HADS) scale, in terms of depression screening and case-finding (Williams et al., 2002a,b). While the 7-item depression subscale of the HADS excludes somatic symptoms that may be confounded by comorbid medical disorders, a meta-analysis of the PHQ-9 showed good operating characteristics across a range of medical
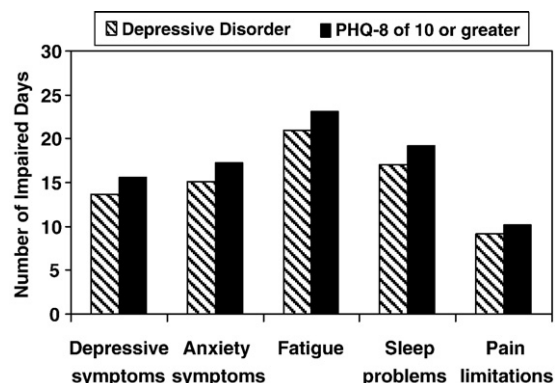

Fig. 2. Number of self-reported days in the past 30 days with depressive symptoms, anxiety symptoms, fatigue, sleep problems, and pain limiting activity. Data available from 13,622 BRFSS respondents in 3 states: Delaware, Hawaii, and Rhode Island.

disorders (Gilbody et al., 2007b). Also, there is not strong empiric data to support the exclusion of somatic symptoms in diagnosing depression in medically ill populations (Koenig et al., 1997). A treatment study in 32 general practices in Scotland found that the PHQ-9 and HADS had similar reliability, discriminant and factorial validity, and sensitivity to change (Cameron et al., 2008). However, the PHQ-9 classified more patients with moderate/severe depression, prompting the authors to conclude that further work is needed to assess the validity of the two scales' severity cut-off bands. A recent validation study in the UK found the PHQ-9 as good as clinician-administered instruments in detecting depression in primary care (Gilbody et al., 2007a).

BRFSS made one important modification to the PHQ-8: respondents were asked to report the number of days they were bothered by each depressive symptom in the past 2 weeks, and then this number was converted to the standard verbal response options of the PHQ-8. These conversions had substantial face validity: 0–1 days is less than "several days" and 7–11 days is "more than half the days." In addition, the training manual for the *Structured Clinical Interview for DSM-IV Axis I Disorders* (SCID) instructs the interviewer that more than 11 days is the cutpoint if respondents insist on explicit guidance for what is meant by "nearly every day in the past 2 weeks" (First et al., 1996). Ideally, the two methods of administering and scoring the PHQ-8 would be tested in a single sample using a structured psychiatric interview as the criterion standard. However, the fact that the population prevalence of current depression detected by the PHQ-8 using the "number of days" response set is comparable to previous epidemiological studies, combined with the strong construct validity demonstrated by our HRQoL findings, provides important preliminary validation data for the "number of days" approach. Of note, a response set inquiring about number of days in the past week is also being tested in several studies using interactive voice recorded PHQ-9 administration. In these studies, 0–1 day is scored as 0 ("not at all"), 2–3 days as 1 ("several days"), 4–5 days as 2 ("more than half the days"), and 6–7 days as 3 ("nearly every day").

Several limitations of our study must be acknowledged. First, BRFSS excludes individuals in institutions and those who cannot participate in a telephone interview (severe mental illness or cognitive impairment, lack of a phone due to low income, etc.). Because these groups may have higher rates of depression, BRFSS may slightly underestimate the true prevalence. Second, although our prevalence and HRQoL impairment data provide strong preliminary evidence for the "number of days" response set as a valid alternative to the standard PHQ response set,

this finding would be strengthened by testing both response sets against a criterion standard in a single sample. Third, despite the similar prevalences, respondent sociodemographics, and HRQoL impairment of current depression as defined by the diagnostic algorithm and the cutpoint approach, there was a moderate degree of non-overlap in the actual groups identified by the two methods. Treatment and other types of longitudinal studies should further compare the prognosis and outcomes of depression defined by the PHQ in these two ways. Fourth, even if a cutpoint approach may suffice for large, epidemiological studies, further inquiry is advisable in the clinical care of individual patients. For example, a PHQ-8 score of 10 or greater in the clinical setting should prompt a careful interview for a number of reasons: to determine if the elevated score represents clinical depression, to consider confounding medical causes, to evaluate for affective disorders that may warrant either more conservative initial treatment than major depression (e.g., minor depression or bereavement) or earlier mental health referral (e.g., bipolar disorder or comorbid substance abuse), and to avoid premature labeling of a person as having depression.

There are a number of depression instruments – varying in length from 2 to 20 or more items – that are generally comparable as measures of potential caseness for the purposes of depression screening (Mitchell and Coyne, 2007; Williams et al., 2002a,b). Moreover, these instruments have similar degrees of association with measures of functional status, disability, and HRQoL (Fleishman and Zuvekas, 2007). Compared with many of these, the PHQ-8 (like the PHQ-9) is equally good as a measure of severity and is superior as a diagnostic measure for DSM-IV depressive disorders. Our data from a representative sample of nearly 200,000 individuals in the United States supports the value of the PHQ-8 in population studies.

Appendix A: Patient Health Questionnaire eight-item depression measure (PHQ-8)
PHQ-8

Over the last 2 weeks, how often have you been bothered by any of the following problems? (Use "✔" to indicate your answer)	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself – or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television.	0	1	2	3
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	0	1	2	3

(For office coding: Total Score _____ = ___ + ___ + ___)

References

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders, 4th Edition. American Psychiatric Association, Washington, DC.

Andresen, E.M., Catlin, T.K., Wyrwich, K.W., Jackson-Thompson, J., 2003. Retest reliability of surveillance questions on health related quality of life. J. Epidemiol. Community Health 57 (5), 339–343.

Bombardier, C.H., Fann, J.R., Temkin, N., Esselman, P.C., Pelzer, E., Keough, M., Dikmen, S., 2006. Posttraumatic stress disorder symptoms during the first six months after traumatic brain injury. J. Neuropsychiatry Clin. Neurosciences 18 (4), 501–508.

Bombardier, C.H., Richards, J.S., Krause, J.S., Tulsky, D., Tate, D.G., 2004. Symptoms of major depression in people with spinal cord injury: implications for screening. Arch. Phys. Med. Rehab. 85 (11), 1749–1756.

Callahan, C.M., Boustani, M.A., Unverzagt, F.W., Austrom, M.G., Damush, T.M., Perkins, A.J., Fultz, B.A., Hui, S.L., Counsell, S.R., Hendrie, H.C., 2006. Effectiveness of collaborative care for older adults with Alzheimer disease in primary care — a randomized controlled trial. JAMA 295 (18), 2148–2157.

Cameron, I.M., Crawford, J.R., Lawton, K., Reid, I.C., 2008. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br. J Gen. Pract. 58, 32–36.

Centers for Disease Control and Prevention, 2005. Behavioral Risk Factor Surveillance System User's Guide. U.S. Department of Health and Human Services. Available at: ftp://ftp.cdc.gov/pub/Data/Brfss/userguide.pdf, Atlanta, GA.

Corson, K., Gerrity, M.S., Dobscha, S.K., 2004. Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item. Am. J. Managed Care 10 (11), 839–845.

Diez-Quevedo, C., Rangil, T., Sanchez-Planell, L., Kroenke, K., Spitzer, R.L., 2001. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. Psychosom. Med 63 (4), 679–686.

Drayer, R.A., Piraino, B., Reynolds, C.F., Houck, P.R., Mazumdar, S., Bernardini, J., Shear, M.K., Rollman, B.L., 2006. Characteristics of depression in hemodialysis patients: symptoms, quality of life and mortality risk. Gen. Hosp. Psychiatry 28 (4), 306–312.

Dwight-Johnson, M., Ell, K., Lee, P.J., 2005. Can collaborative care address the needs of low-income Latinas with comorbid depression and cancer? Results from a randomized pilot study. Psychosomatics 46 (3), 224–232.

Fann, J.R., Bombardier, C.H., Dikmer, S., Esselman, P., Warms, C.A., Pelzer, E., Rau, H., Temkin, N., 2005. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. J. Head Trauma Rehab. 20 (6), 501–511.

First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B., 1996. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID). American Psychiatric Press Inc.

Fleishman, J.A., Zuvekas, S.H., 2007. Global self-rated mental health: associations with other mental health measures and with role functioning. Med. Care 45 (7), 602–609.

Gilbody, S., Richards, D., Barkham, M., 2007a. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. Br. J Gen. Pract. 57, 650–652.

Gilbody, S., Richards, D., Brealey, S., Hewitt, C., 2007b. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J. Gen. Intern. Med. 22, 1596–1602.

Glasgow, R.E., Nutting, P.A., King, D.K., Nelson, C.C., Cutter, G., Gaglio, B., Rahm, A.K., Whitesides, H., Amthauer, H., 2004. A practical randomized trial to improve diabetes care. J. Gen. Intern. Med. 19 (12), 1167–1174.

Hawthorne, G., Cheok, F., Goldney, R., Fisher, L., 2003. The excess cost of depression in South Australia: a population-based study. Aust. N. Zealand J. Psychiatry 37 (3), 362–373.

Helzer, J.E., Kraemer, H.C., Krueger, R.F., 2006. The feasibility and need for dimensional psychiatric diagnoses. Psychol. Med. 36, 1671–1680.

Holtzman, D., 2004. The Behavioral Risk Factor Surveillance System. In: Blumenthal, D.S., DiClemente, R.J. (Eds.), Community-based Health Research Issues and Methods. Springer, New York, pp. 115–131.

Holzapfel, N., Zugck, C., Muller-Tasch, T., Lowe, B., Wild, B., Schellberg, D., Nelles, M., Herzog, W., Junger, J., 2007. Routine screening for depression and quality of life in outpatients with congestive heart failure. Psychosomatics 48 (2), 112–116.

Huang, F.Y., Chung, H., Kroenke, K., Delucchi, K.L., Spitzer, R.L., 2006a. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. J. Gen. Intern. Med. 21 (6), 547–552.

Huang, F.Y., Chung, H., Kroenke, K., Spitzer, R.L., 2006b. Racial and ethnic differences in the relationship between depression severity and functional status. Psychiatr. Serv. 57 (4), 498–503.

Justice, A.C., McGinnis, K.A., Atkinson, J.H., Heaton, R.K., Young, C., Sadek, J., Madenwald, T., Becker, J.T., Conigliaro, J., Brown, S.T., Rimland, D., Crystal, S., Simberkoff, M., 2004. Psychiatric and neurocognitive disorders among HIV-positive and negative veterans in care: Veterans Aging Cohort Five-Site Study. AIDS 18, S49–S59.

Katon, W.J., Simon, G., Russo, J., Von Korff, M., Lin, E.H.B., Ludman, E., Ciechanowski, P., Bush, T., 2004. Quality of depression care in a population-based sample of patients with diabetes and major depression. Med. Care 42 (12), 1222–1229.

Kendall, R., Jablensky, A., 2003. Distinguishing between the validity and utility of psychiatric diagnoses. Am. J. Psychiatry 160, 4–12.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S., 2003. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 289 (23), 3095–3105.

Kessler, R.C., McGonagle, K.A., Zhao, S., Nelson, C.B., Hughes, M., Eshelman, S., Wittchen, H., Kendler, K.S., 1994. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Arch. Gen. Psychiatry 51 (1), 8–19.

Klapow, J., Kroenke, K., Horton, T., Schmidt, S., Spitzer, R., Williams, J.B., 2002. Psychological disorders and distress in older primary care patients: a comparison of older and younger samples. Psychosom. Med. 64 (4), 635–643.

Koenig, H.G., George, L.K., Peterson, B.L., Pieper, C.F., 1997. Depression in medically ill hospitalized older adults: prevalence, characteristics, and course of symptoms according to six diagnostic schemes. Am. J. Psychiatry 154, 1376–1383.

Kroenke, K., Spitzer, R.L., 2002. The PHQ-9: a new depression and diagnostic severity measure. Psychiatric Ann. 32, 509–521.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613.

Lee, P.W., Schulberg, H.C., Raue, P.J., Kroenke, K., 2007. Concordance between the PHQ-9 and the HSCL-20 in depressed primary care patients. J Affect. Disord. 99 (1–3), 139–145.

Lowe, B., Kroenke, K., Herzog, W., Grafe, K., 2004a. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J Affect. Disord. 81 (1), 61–66.

Lowe, B., Unutzer, J., Callahan, C.M., Perkins, A.J., Kroenke, K., 2004b. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med. Care 42 (12), 1194–1201.

Lowe, B., Willand, L., Eich, W., Zipfel, S., Ho, A.D., Herzog, W., Fiehn, C., 2004c. Psychiatric comorbidity and work disability in patients with inflammatory rheumatic diseases. Psychosom. Med. 66 (3), 395–402.

Maizels, M., Smitherman, T.A., Penzien, D.B., 2006. A review of screening tools for psychiatric comorbidity in headache patients. Headache 46, S98–S109.

Martin, A., Rief, W., Klaiberg, A., Braehler, E., 2006. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen. Hosp. Psychiatry 28 (1), 71–77.

Mielenz, T., Jackson, E., Currey, S., DeVellis, R., Callahan, L.F., 2006. Psychometric properties of the Centers for Disease Control and Prevention Health-Related Quality of Life (CDC HRQOL) items in adults with arthritis. Health Qual. Life Outcomes. 4, 66.

Mitchell, A.J., Coyne, J.C., 2007. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. Br. J Gen. Pract. 57 (535), 144–151.

Mokdad, A.H., Stroup, D.F., Giles, W.H., 2003. Public health surveillance for behavioral risk factors in a changing environment. Recommendations from the Behavioral Risk Factor Surveillance Team. MMWR Recomm. Rep. 52 (RR-9), 1–12.

Moriarty, D.G., Zack, M.M., Kobau, R., 2003. The Centers for Disease Control and Prevention's Healthy Days Measures — population tracking of perceived physical and mental health over time. Health Qual. Life Outcomes. 1 (1), 37.

Persoons, P., Luyckx, K., Desloovere, C., Vandenberghe, J., Fischler, B., 2003. Anxiety and mood disorders in otorhinolaryngology outpatients presenting with dizziness: validation of the self-administered PRIME-MD Patient Health Questionnaire and epidemiology. Gen. Hosp. Psychiatry 25 (5), 316–323.

Persoons, P., Luyckx, K., Fischler, B., 2001. Psychiatric diagnoses in gastroenterology: validation of a self-report instrument (PRIME-MD Patient Health Questionnaire), epidemiology and recognition. Gastroenterology 120 (5), 615.

Picardi, A., Amerio, P., Baliva, G., Barbieri, C., Teofoli, P., Bolli, S., Salvatori, V., Mazzotti, E., Pasquini, P., Abeni, D., 2004. Recognition of depressive and anxiety disorders in dermatological outpatients. Acta Dermato-Venereologica 84 (3), 213–217.

Pinto-Meza, A., Serrano-Blanco, A., Penarrubia, M.T., Blanco, E., Haro, J. M., 2005. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? J. Gen. Intern. Med. 20 (8), 738–742.

Rief, W., Nanke, A., Klaiberg, A., Braehler, E., 2004. Base rates for panic and depression according to the Brief Patient Health Questionnaire: a population-based study. J. Affect. Disord. 82 (2), 271–276.

Robins, L.N., Regier, D.A., 1991. Psychiatric Disorders in America: The Epidemiologic Catchment Area Study. Free Press, New York.

Rosemann, T., Backenstrass, M., Joest, K., Rosemann, A., Szecsenyi, J., Laux, G., 2007. Predictors of depression in a sample of 1,021 primary care patients with osteoarthritis. Arthritis & Rheumatism-Arthritis Care Res. 57 (3), 415–422.

Ruo, B., Rumsfeld, J.S., Hlatky, M.A., Liu, H., Browner, W.S., Whooley, M.A., 2003. Depressive symptoms and health-related quality of life: the Heart and Soul Study. JAMA 290 (2), 215–221.

Scholle, S.H., Haskett, R.F., Hanusa, B.H., Pincus, H.A., Kupfer, D.J., 2003. Addressing depression in obstetrics/gynecology practice. Gen. Hosp. Psychiatry 25 (2), 83–90.

Smith, L., Gilhooly, K., 2006. Regression versus fast and frugal models of decision-making: the case of prescribing for depression. Appl. Cogn. Psychol. 20, 265–274.

Spitzer, R.L., Williams, J.B.W., Kroenke, K., Hornyak, R., McMurray, J., Heartwell, S.F., for the Patient Health Questionnaire Obstetrics Gynecology Study Group, 2000. Validity and utility of the Patient Health Questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. Am. J. Obstet. Gynecol. 183 (3), 759–769.

Stewart, W.F., Ricci, J.A., Chee, E., Hahn, S.R., Morganstein, D., 2003. Cost of lost productive work time among US workers with depression. JAMA 289 (23), 3135–3144.

Tietjen, G.E., Brandes, J.L., Digre, K.B., Baggaley, S., Martin, V., Recober, A., Geweke, L.O., Hafeez, F., Aurora, S.K., Herial, N.A., Utley, C., Khuder, S.A., 2007. High prevalence of somatic symptoms and depression in women with disabling chronic headache. Neurology 68 (2), 134–140.

Turner, J.A., Dworkin, S.F., 2004. Screening for psychosocial risk factors in patients with chronic orofacial pain — recent advances. J. Am. Dental Assoc. 135 (8), 1119–1125.

Turvey, C.L., Willyard, D., Hickman, D.H., Klein, D.M., Kukoyi, O., 2007. Telehealth screen for depression in a chronic illness care management program. Telemedicine J. E-Health 13 (1), 51–56.

Williams Jr., J.W., Pignone, M., Ramirez, G., Perez, S.C., 2002a. Identifying depression in primary care: a literature synthesis of case-finding instruments. Gen. Hosp. Psychiatry 24 (4), 225–237.

Williams, J.W.J., Noel, P.H., Cordes, J.A., Ramirez, G., Pignone, M., 2002b. Is this patient clinically depressed? JAMA 287, 1160–1170.

Williams, L.S., Brizendine, E.J., Plue, L., Bakas, T., Tu, W.Z., Hendrie, H., Kroenke, K., 2005. Performance of the PHQ-9 as a screening tool for depression after stroke. Stroke 36 (3), 635–638.

Williams, L.S., Jones, W.J., Shen, J., Robinson, R.L., Kroenke, K., 2004. Outcomes of newly referred neurology outpatients with depression and pain. Neurology 63 (4), 674–677.

World Health Organization, 2002. The World Health Report 2002: Reducing Risks, Promoting Healthy Life. World Health Organization, Geneva, Switzerland.