ORIGINAL ARTICLES

Using the Patient Health Questionnaire-9 to Measure Depression among Racially and Ethnically Diverse Primary Care Patients

Frederick Y. Huang, MD,[1] Henry Chung, MD,[2,3] Kurt Kroenke, MD,[4,5] Kevin L. Delucchi, PhD,[1] Robert L. Spitzer, MD[6,7]

[1]Department of Psychiatry, University of California, San Francisco, CA, USA; [2]Charles B. Wang Community Health Center, New York, NY, USA; [3]New York University, New York, NY, USA; [4]Indiana University School of Medicine, Indianapolis, IN, USA; [5]Regenstrief Institute, Inc., Indianapolis, IN, USA; [6]Biometrics Research Department, New York Psychiatric Institute, New York, NY, USA; [7]Department of Psychiatry, Columbia University, New York, NY, USA.

OBJECTIVE: The Patient Health Questionnaire depression scale (PHQ-9) is a well-validated, Diagnostic and Statistical Manual of Mental Disorders— Fourth Edition (DSM-IV) criterion-based measure for diagnosing depression, assessing severity and monitoring treatment response. The performance of most depression scales including the PHQ-9, however, has not been rigorously evaluated in different racial/ethnic populations. Therefore, we compared the factor structure of the PHQ-9 between different racial/ethnic groups as well as the rates of endorsement and differential item functioning (DIF) of the 9 items of the PHQ-9. The presence of DIF would indicate that responses to an individual item differ significantly between groups, controlling for the level of depression.

MEASUREMENTS: A combined dataset from 2 separate studies of 5,053 primary care patients including non-Hispanic white ($n=2,520$), African American ($n=598$), Chinese American ($n=941$), and Latino ($n=974$) patients was used for our analysis. Exploratory principal components factor analysis was used to derive the factor structure of the PHQ-9 in each of the 4 racial/ethnic groups. A generalized Mantel-Haenszel statistic was used to test for DIF.

RESULTS: One main factor that included all PHQ-9 items was found in each racial/ethnic group with α coefficients ranging from 0.79 to 0.89. Although endorsement rates of individual items were generally similar among the 4 groups, evidence of DIF was found for some items.

CONCLUSIONS: Our analyses indicate that in African American, Chinese American, Latino, and non-Hispanic white patient groups the PHQ-9 measures a common concept of depression and can be effective for the detection and monitoring of depression in these diverse populations.

M ajor depression is 1 of the most common psychiatric disorders. The National Comorbidity Survey Replication estimates the lifetime prevalence of major depressive disorder to be 16% among adults in the United States.[1] Moreover, depressive illness is projected to have significant public health and economic costs: major depression is expected to be the

second leading cause of death and disability and to impose the greatest burden of ill health worldwide by 2020.[2] Depression, however, is frequently unrecognized and not treated.[3] The U.S. Preventive Services Task Force (USPSTF) has therefore recommended systematic screening of depression in clinical settings with appropriate systems in place to ensure effective treatment and follow-up.[4]

The PHQ-9 is the depression module of the self-administered version of the PRIME-MD diagnostic instrument, called the Patient Health Questionnaire (PHQ).[5] The PHQ-9 is an instrument whose 9 items are based on the DSM-IV diagnostic criteria. Each of the 9 items can be scored from 0 (not at all) to 3 (nearly every day). Its validity and reliability as a diagnostic measure as well as its utility in assessing depression severity and monitoring treatment response are well-established.[5–12]

No studies, however, have yet examined the factor structure and differential functioning of the PHQ-9 in different racial/ethnic minority groups. In fact, relatively few studies have systematically investigated racial/ethnic differences in the function of other depression screening instruments.[13–21] Assessing the performance of the PHQ-9 and other depression measures in different racial/ethnic groups is of growing concern because of the increasing diversity in the United States. According to the U.S. Census Bureau, Latinos are the largest racial/ethnic minority group in the country, numbering more than 35 million, whereas African Americans have historically constituted a numerically important minority group in the United States, currently numbering 34 million persons.[22] Asian Americans are a fast-growing racial/ethnic group, increasing 44% from 1990 to 2000.[23] It is projected that by the year 2020 the Asian American population will reach approximately 20 million.[24] This growing diversity in the United States makes the need for diagnostic instruments that can provide accurate, clinically relevant information for depression across racial/ethnic groups an urgent public health priority, especially in light of the USPSTF recommendation for screening of depression in routine care.

In this paper, data from the original PHQ Primary Care and Obstetrics/Gynecology validation studies were combined with PHQ-9 screening data collected from Chinese American

patients attending a large urban primary care clinic. We compared the factor structure as well as the individual items of the PHQ-9 in 4 racial/ethnic subgroups: African American, Chinese American, Latino, and non-Hispanic whites.

METHODS

Study Design

Data from the PHQ Primary Care and Obstetrics/Gynecology Studies were combined with PHQ-9 depression screening data collected from the Charles B. Wang Community Health Center (CBWCHC), a community health center serving mostly low-income Chinese Americans and affiliated with the NYU School of Medicine Center for the Study of Asian American Health.

The process of subject selection, administration of instruments, and data collection of the PHQ Primary Care and Obstetrics/Gynecology Studies are explained in detail in prior publications.[7,8] Briefly, from May 1997 to November 1998, 3,000 primary care patients (1,422 from 5 general internal medicine clinics and 1,578 from 3 family practice clinics) participated in the PHQ Primary Care Study. From May 1997 to March 1999, 3,000 patients from 7 obstetrics-gynecology outpatient sites participated in the PHQ Obstetrics/Gynecology Studies. Study subjects in both studies were selected by 2 selection methods to minimize sampling bias: consecutive patients for a given clinic session or every *n*th patient until the intended quota for that session was achieved. The patients in both studies, all of whom were 18 years or older, completed the PHQ-9 before seeing their physician.

The Chinese American sample was composed of 3,417 primary care patients at the CBWCHC in New York City who were scheduled for their annual physical examination and pre-screened with a survey: They were asked whether in the past 2 weeks, they had experienced anhedonia, depressed mood, low energy, or insomnia. Patients who endorsed at least 1 of these items were then administered the PHQ-9. This resulted in 973 individuals at the CBWCHC who filled out the PHQ-9.

As subjects from the PHQ Studies were not screened before PHQ-9 administration, we included for analysis only those subjects who, similar to the CBWCHC sample, endorsed positively at least 1 of the PHQ-9 items of anhedonia, depressed mood, insomnia, or low energy. This resulted in a total of 5,033 subjects from the PHQ Primary Care Study (*n*=1,964), PHQ Obstetrics/Gynecology Study (*n*=2,128), and CBWCHC site (*n*=941).

Measure

Depression Scale of the PHQ-9. At all sites, including CBWCHC, the PHQ-9 was administered immediately before the physician encounter. In most cases, the PHQ-9 was self-completed by the patient in written form. If, however, the patient was unable to read or needed other assistance, the PHQ-9 was read to the patient by a care manager. Patients either monolingual Spanish-speaking or Chinese-speaking were administered the Spanish or Chinese versions of the PHQ-9 respectively. Both of these versions of the PHQ-9 were translated into the non-English language and back translated into English by bilingual translators. Consistent with the process for translation of survey instruments outlined in prior studies,[25] translation and back translation of the Spanish and Chinese language versions of the PHQ-9 was repeated until the trans-

lators felt the non-English versions corresponded closely with the English version.

Statistical Analysis

Sociodemographic differences in gender, age, marital status, and language ability were compared between racial/ethnic groups. Mean differences in total PHQ-9 scores and rates of depression severity were also compared between the different racial/ethnic groups. Endorsement of the 9 individual items of the PHQ-9 was compared using mean scores on the individual items.

Differential item functioning (DIF) is a method for determining if an item on a scale, controlling for the latent factor the scale measures, is endorsed more often by one subgroup than another. Graphically, the DIF can be conceptualized as detecting the presence of distinct s-shaped item-response curves for subgroups as opposed to a single s-shaped curve. Differential item functioning is a concept derived from item response theory (IRT), which is a modern measurement theory that focuses on comparing responses to individual items in a scale rather than simply the aggregate scale score.

In our analysis, the individual items of the PHQ-9 were analyzed for DIF by comparing the endorsement of items between different groups controlling for the level of depression as indexed, in this case, by the total PHQ-9 score. The test in this analysis is an extension of the Mantel-Haenszel statistic, which tests for the equality of odds ratios across several strata.[26] Three separate analyses were conducted, each comparing the non-Hispanic white group (referent group) to 1 of the ethnic groups.

A limitation of the Mantel-Haenszel statistic is that it cannot control for additional covariates beyond depression level. To control for these effects we estimated and tested a Multiple Indicators Multiple Causes (MIMIC) model, which is a test of DIF that included age, sex, and English language ability as covariates.[27]

The factor structure of the PHQ-9 within each ethnic group was examined using principal component analysis. The basis of factor analysis lies in the concept that the responses to the items in a scale are the result of an underlying, latent dimension or set of dimensions—in this case, the latent dimension is depression. If the set of items operates in a consistent fashion across groups, the underlying factor structure will be the same and should display a large eigenvalue. In our analysis, components with an eigenvalue of 1.0 or greater were retained and rotated to a varimax solution. The Kaiser criterion was used to select the number of factors and the oblique solution was employed.[28] Internal reliability of the factor structure for each racial/ethnic group was assessed by calculating the Cronbach's α coefficient for each factor.

Chi-square tests were used for categorical data and analysis of variance for continuous data. Bonferroni's correction was used to adjust for multiple comparisons. All statistical analyses were done using SPSS for Windows 9.0 (SPSS Inc., Chicago, IL), SAS version 9.13 (SAS Institute, Cary, NC), and Mplus.[29]

RESULTS

Patient Characteristics

Table 1 summarizes the baseline characteristics of the patient samples divided by racial/ethnic group, illustrating some significant group differences. Latino patients were the youngest

Table 1. Baseline Patient Characteristics

	African American	Chinese American	Latino	Non-Hispanic White
Mean ± SD age, y*	38.6 ± 15.2	43.1 ± 13.6	29.4 ± 10.0	39.9 ± 16.3
Female (%)(*n*)[†]	85.3 (510)	53.3 (501)	97.8 (952)	80.6 (2,031)
Married (%)(*n*)[†]	33.2 (199)	75.2 (708)	56.0 (545)	50.5 (1,273)
Monolingual non-English-speaking (%)(*n*)[†]	1.2 (7)	97.4 (917)	73.6 (717)	0.3 (8)

*P<.001 (ANOVA), post-hoc tests show significant differences between all groups (P<.001) except African American and non-Hispanic whites.
[†]P<.001 (χ^2 test).

(29.4 years) as well as the most likely to be female (97.8%). The Chinese American group was the oldest (43.1 years) and most likely to be married (75.2%). Both the Latino and Chinese American groups were most likely to be non-English-speaking with 73.6% of the Latino group being monolingual Spanish-speaking and 97.4% of the Chinese American group being monolingual Chinese speaking.

Mean PHQ-9 Scores and Distribution of Depression Severity Levels

The comparison of mean scores is presented in Figure 1. No significant differences were found between the 4 groups in the mean PHQ-9 score. The mean scores ranged from a low of 6.0 (African Americans) to a high of 6.5 (Chinese Americans). A PHQ-9 score greater than or equal to 10 typically represents clinically significant depression symptoms with an 88% sensitivity and specificity for the diagnosis of depression.[5] There were significant differences among the racial/ethnic groups (χ^2=21.16, df=3, P<.001) in the proportion of patients exceeding this threshold, ranging from 15.2% (Chinese Americans) to 21.8% (non-Hispanic whites). In addition, cases of depression were categorized into different levels of severity: moderate,[10–14] moderately severe,[15–19] and severe.[5,20–27] Chinese American subjects had the lowest frequency (7.3%) of
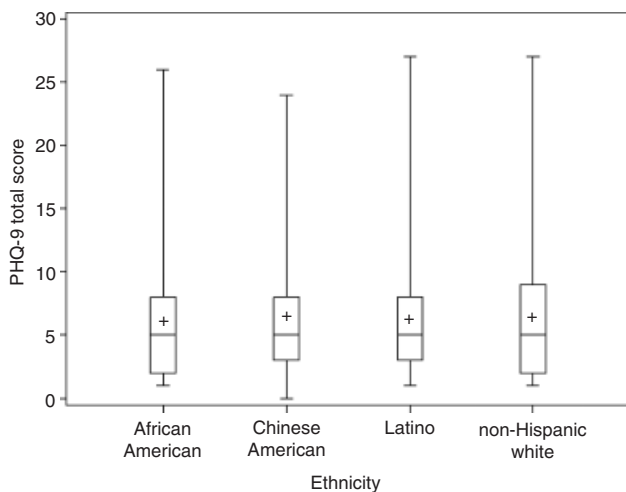

FIGURE 1. Patient Health Questionnaire depression scale (PHQ-9) total score distribution by ethnic group. The bottom and top ends of the boxes denote the 25th and 75th percentiles, respectively, of PHQ-9 total scores. The lines within each of the boxes denote the median score and the + signs denote the mean score. The lines stemming from each of the boxes display the range of remaining values.

moderate levels of depression (χ^2=15.05, df=3, P<.005). No significant group differences were seen in the rates of moderately severe or severe levels of depression.

Separate analyses of male and female patients revealed that Chinese American men had especially low rates of depression compared with all other groups. This was in contrast with Chinese American women who had comparative levels of depression to all other racial/ethnic groups: 11.8% of Chinese American men (compared with 18.1% of Chinese American women) had PHQ-9 scores of 10 or higher, and 4.8% of Chinese American men (compared with 9.6% of Chinese American women) had PHQ-9 scores between 10 and 14.

Factor Analysis

Results of exploratory factor analyses are shown in Table 2. In each of the 4 groups, a single factor that included all 9 items of the PHQ-9 was extracted. The eigenvalue on this single factor ranged from 3.50 (Chinese American) to 4.42 (non-Hispanic white), and the variance explained by this single factor ranged from 38.9% (Chinese American) to 49.1% (non-Hispanic white). Internal consistency reliability (Cronbach's α) of the PHQ-9 was 0.80, 0.79, 0.80, and 0.86 in African Americans, Chinese Americans, Latinos, and non-Hispanic whites, respectively.

Individual Item Analysis

Comparisons between the different racial/ethnic groups of mean scores of each of the 9 items of the PHQ-9 are shown in Figure 2. Individual items of the PHQ-9 are scored on a scale of 0 to 3 for symptoms occurring in the past 2 weeks. Zero corresponds with the symptom occurring "not at all," 1 with "several days," 2 with "more than half the days," and 3 with "nearly every day." Overall, the 2 items that were endorsed most frequently in all groups were abnormalities in sleep and low energy. Scores for these 2 items ranged from 0.96 to 1.37 for abnormalities in sleep and 1.24 to 1.41 for low energy.

Three items—depressed mood, decreased concentration, and thoughts of death or self-harm—were not significantly different between groups. Chinese American and Latino subjects showed different patterns of endorsement in the other individual PHQ-9 items. Chinese Americans endorsed psychomotor abnormalities at a rate more than double the other groups: 0.8 compared with 0.24 to 0.35 (F=104.99, df=3, P<.001). They endorsed abnormalities of appetite at a rate less than half the other groups: 0.4 compared with 0.85 to 0.95 (F=64.10, df=3, P<.001). Chinese Americans also had significantly higher mean scores in abnormalities in sleep (F=25.71, df=3, P<.001). Latinos had significantly higher mean scores of anhedonia: 0.89 compared with 0.56 to 0.67 (F=23.63,

Table 2. Major Factors and Loadings of Patient Health Questionnaire-9 Items

	Factor 1			
	African American, N=598	Chinese American, N=941	Latino, N=974	Non-Hispanic, White N=2,520
Anhedonia	0.694	0.682	0.639	0.748
Depressed mood	0.708	0.755	0.728	0.796
Sleep problems	0.551	0.425	0.567	0.604
Low energy	0.533	0.474	0.632	0.617
Appetite change	0.592	0.572	0.564	0.687
Low self-esteem	0.743	0.667	0.648	0.785
Concentration difficulties	0.662	0.634	0.689	0.756
Psychomotor agitation or retardation	0.624	0.767	0.623	0.659
Suicidal ideation	0.556	0.544	0.548	0.624
Eigenvalues	3.61	3.50	3.56	4.42
Variance explained (%)	40.1	38.9	39.6	49.1

df=3, $P<.001$). In comparison with Chinese Americans and non-Hispanic whites but not African Americans, Latinos had lower mean scores of abnormalities in sleep ($F=25.71$, df=3, $P<.001$), low energy ($F=7.85$, df=3, $P<.001$), and guilt ($F=11.61$, df=3, $P<.001$).

Table 3 shows that many of the same individual items that had statistically different mean scores in Chinese American and Latino groups also showed evidence of DIF when compared with the non-Hispanic white group. For the Chinese American group, sleep, appetite, and psychomotor changes showed DIF. Anhedonia also had a significant Mantel-Haenzel statistic in the Chinese American group, but after controlling for covariates of age, sex, and English-language ability in the MIMIC model test, this item no longer showed a significant level of DIF. In the Latino group, anhedonia, sleep changes, appetite changes, and guilt showed evidence of DIF. The MIMIC model test indicated that the depressed mood and low energy items did not have significant DIF when controlling for sociodemographic factors. No significant DIF was found for any of the individual items in the African American group.
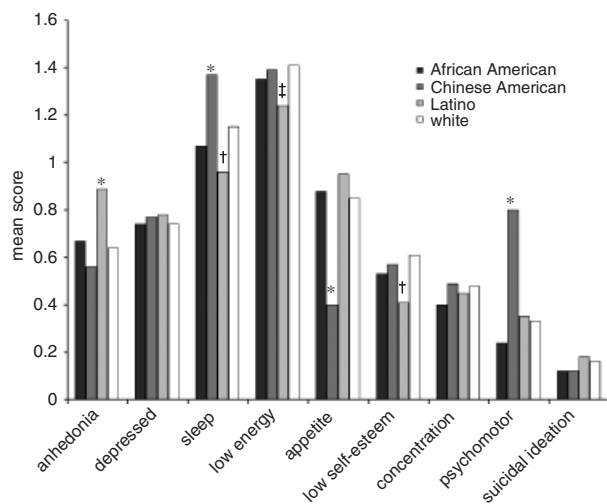

FIGURE 2. Mean scores of individual Patient Health Questionnaire depression scale items. *$P<.001$, compared with all other ethnic groups; †$P<.001$, except compared with African Americans; ‡$P=.005$, except compared with African Americans.

DISCUSSION

The results of this analysis demonstrate that the PHQ-9 total score functions fundamentally the same in subjects from 4 of the largest racial/ethnic groups in the United States. The similar mean scores and factor structure of the PHQ-9 in the different groups—even while the vast majority of Chinese Americans and Latinos in this analysis completed the PHQ9 in a language other than English—suggests that it can be used without adjustment in diverse populations. These findings also support the idea that the DSM-IV criteria for major depression are common to individuals of all cultures.

Mean PHQ-9 scores were similar amongst the different racial/ethnic groups. Unlike what has been seen in previous studies, Latinos in our study did not have higher mean PHQ-9 scores compared with the other racial/ethnic groups.[30,31] Our findings imply that there is no need to adjust PHQ-9 threshold scores for depression in patients from Latino backgrounds.[32] One possible reason for this difference between our study and past research is that prior studies used measures such as the CES-D or Beck Depression Inventory that were not strictly criterion-based, unlike the PHQ-9.

Chinese Americans also did not have significantly different mean scores than individuals from other groups, but a smaller proportion of Chinese Americans had a clinically significant level of depressive symptoms as indicated by a score of 10 or higher. Gender stratification revealed much lower rates of depressive symptoms among Chinese American men while Chinese American women had rates of depression statistically indistinguishable from other racial/ethnic groups. These gender differences are consistent with prior studies[33] as well as epidemiological studies of Chinese Americans that showed acculturated Chinese American women have twice the likelihood of lifetime depressive episodes as Chinese American men.[34] Future studies could explore further this possibility that gender mediates the effect of acculturation on endorsement of depressive symptoms.

This study is the first to examine the factor structure of the PHQ-9. The fact that all 9 of these items load onto a single factor suggests that the PHQ-9 is measuring a coherent, unitary concept of major depressive disorder based on the DSM-IV criteria. Furthermore, the fact that this single factor comprising all 9 items is seen in all 4 racial/ethnic groups suggests that the core features of depression are common in

Table 3. Test and Associated *P*-Values for Differential Item Functioning

PHQ-9 Items	African American		Chinese American		Latino	
	Mantel-Haenszel statistic	*P*-Value	Mantel-Haenszel statistic	*P*-Value	Mantel-Haenszel statistic	*P*-Value
Anhedonia	7.67	.0056	8.42	**.0037**	116.92	**<.0001**
Depressed mood	2.98	.0845	1.17	.28	7.86	**.0051**
Sleep problems	1.9	.1686	24.87	**<.0001**	41.71	**<.0001**
Low energy	0.59	.4443	6.57	.0104	38.9	**<.0001**
Appetite changes	3.69	.0546	267.52	**<.0001**	15.01	**<.0001**
Low self-esteem	0.55	.4593	3.1	.0781	52.06	**<.0001**
Concentration difficulties	2.72	.0989	0.4	.5287	0.01	.9417
Psychomotor agitation or retardation	2.83	.0923	382.2	**<.0001**	6.77	.0093
Suicidal ideation	0.06	.8049	0.91	.3409	6.66	.0099

Non-Hispanic white was the referent group. All tests 1 degree-of-freedom.
P-values in bold type significant at P<.05/9 =.0055.

these groups. Although there may be some differences in the expression of individual symptoms across racial/ethnic groups, these differences are relatively minor. In our analysis, there was consistency in the core features of depression across racial/ethnic groups, similar to what was found in past cross-national studies.[35,36] This is illustrated by the fact that between all 4 groups there was no significant difference in mean scores of the individual item of depressed mood.

Our finding that Chinese Americans have higher rates of endorsement of psychomotor abnormalities and sleep is consistent with previous studies showing that Asian subjects are more likely to have somatic symptoms.[13,37,38] Conversely, we have not seen any previous literature that parallels our finding that Asian subjects had much fewer symptoms of appetite changes. Future studies of depressive symptomatology that include Chinese Americans should help to confirm or refute this finding.

Our finding that Latino subjects had a higher endorsement of anhedonia on individual item analysis shows similarities with studies of other depression screening instruments. These previous studies have found that Latinos report positive emotional states less often than non-Hispanic whites.[18,31] Similar to our study, prior research has also shown that after controlling for sociodemographic factors there was no difference found in the endorsement of symptoms of negative affect or somatic disturbance between Latinos and non-Hispanic white populations.[30]

Our study has several limitations. It is possible that the exclusion from analysis of those African American, Latino, and non-Hispanic white subjects who did not endorse at least a screening item (depressed mood, anhedonia, insomnia, or low energy) may have led to reduced variance in the response to the PHQ-9 overall. This reduced variance may account for the lack of difference we saw in the function and dimensionality of the PHQ-9 between groups. Finally, while the construct validity of the PHQ-9 has been examined in a separate analysis of these data,[39] convergent validity of the PHQ-9 in these different ethnic groups with an independent criterion standard such as the SCID would also be a valuable future study.

In light of the growing diversity of the U.S. population and the increasingly recognized importance of screening for depression in clinical care, the need for an efficient depression screening instrument that can be used in disparate groups is critical. Our study suggests that the PHQ-9 can be used without ad-

justment in different racial/ethnic groups and be a useful tool to help meet the mental health care needs of diverse populations.

REFERENCES

1. **Kessler RC, Berglund P, Demler O, et al.** The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA. 2003;289:3095–105.
2. **Murray CJ, Lopez AD.** Global mortality, disability, and the contribution of risk factors: global burden of disease study. Lancet. 1997;349:1436–42.
3. **Simon GE, VonKorff M.** Recognition, management, and outcomes of depression in primary care. Arch Fam Med. 1995;4:99–105.
4. **Pignone MP, Gaynes BN, Rushton JL, et al.** Screening for depression in adults: a summary of the evidence for the US. Preventive Services Task Force. Ann Intern Med. 2002;136:765–76.
5. **Kroenke K, Spitzer RL, Williams JB.** The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16:606–13.
6. **Spitzer RL, Kroenke K, Williams JB.** Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA. 1999;282:1737–44.
7. **Spitzer RL, Williams JB, Kroenke K, et al.** Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD patient health questionnaire obstetrics-gynecology study. Am J Obstet Gynecol. 2000;183:759–69.
8. **Lowe B, Unutzer J, Callahan CM, et al.** Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care. 2004;42:1194–201.
9. **Williams JW Jr., Pignone M, Ramirez G, et al.** Identifying depression in primary care: a literature synthesis of case-finding instruments. Gen Hosp Psychiatry. 2002;24:225–37.
10. **Williams JW Jr., Noel PH, Cordes JA, et al.** Is this patient clinically depressed? JAMA. 2002;287:1160–70.
11. **Lowe B, Spitzer RL, Grafe K, et al.** Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. J Affect Disord. 2004;78:131–40.
12. MacArthur Foundation Depression in Primary Care Initiative, www.depression-primarycare.org.
13. **Ying YW.** Depressive symptomatology among Chinese-Americans as measured by the CES-D. J Clin Psychol. 1988;44:739–46.
14. **Posner SF, Stewart AL, Marin G, et al.** Factor variability of the Center for Epidemiological Studies Depression Scale (CES-D) among urban Latinos. Ethn Health. 2001;6:137–44.

15. **Penley JA, Wiebe JS, Nwosu A.** Psychometric properties of the Spanish Beck Depression Inventory-II in a medical sample. Psychol Assess. 2003;15:569–77.

16. **Roberts RE.** Manifestation of depressive symptoms among adolescents. A comparison of Mexican Americans with the majority and other minority populations. J Nerv Mental Dis. 1992;180:627–33.

17. **Hsiao-Rei Hicks M.** Validity of the CIDI Probe flow chart for depression in Chinese American Women. Transcultural Psychiatry. 2002;39:434–51.

18. **Azocar F, Arean P, Miranda J, et al.** Differential item functioning in a Spanish translation of the Beck Depression Inventory. J Clin Psychol. 2001;57:355–65.

19. **Brown C, Schulberg HC, Madonia MJ.** Clinical presentations of major depression by African Americans and whites in primary medical care practice. J Affect Disord. 1996;41:181–91.

20. **Cho MJ, Moscicki EK, Narrow WE, et al.** Concordance between two measures of depression in the Hispanic Health and Nutrition Examination Survey. Soc Psychiatry Psychiatr Epidemiol. 1993;28:156–63.

21. **Zung WW, MacDonald J, Zung EM.** Prevalence of clinically significant depressive symptoms in black and white patients in family practice settings. Am J Psychiatry. 1988;145:882–3.

22. **U.S. Census Bureau**. Overview of Race and Hispanic Origin: Census 2000 Brief. Washington, DC: U.S. Census Bureau; 2002.

23. **U.S. Census Bureau**. The Asian and Pacific Islander Population in the United States: March 2000 (Update). Washington, DC: U.S. Census Bureau; 2001.

24. **U.S. Census Bureau**. Projections of the resident population by race, Hispanic origin, and nativity: Middle series, 2001 to 2005. Washington, DC: U.S. Census Bureau; 2000.

25. **Shumaker SA, Berzon RA.** The International Assessment of Health-related Quality of Life: Theory, Translation, Measurement, and Analysis. Oxford: Rapid Communications; 1995.

26. **Mantel N.** Chi-Square tests with one degree of freedom; extensions of the mantel-haenszel procedure. J Am Stat Assoc. 1963;58:690–700.

27. **Muthen BO, Kao CF, Burstein L.** Instructionally sensitive psychometrics: application of a new IRT-based detection technique to mathematics achievement test items. J Educ Meas. 1991;28:1–22.

28. **Kaiser HF.** The application of electronic computers to factor analysis. Educ Psychol Meas. 1960;20.

29. **Muthen LK, Muthen BO.** Mplus User's Guide. Los Angeles, CA: Muthen & Muthen; 1998–2004.

30. **Aneshensel CS, Clark VA, Frerichs RR.** Race, ethnicity, and depression: a confirmatory analysis. J Pers Soc Psychol. 1983;44:385–98.

31. **Golding JM, Lipton RI.** Depressed mood and major depressive disorder in two ethnic groups. J Psychiatr Res. 1990;24:65–82.

32. **Dressler WM, Badger LW.** Epidemiology of depressive symptoms in black communities. A comparative analysis. J Nerv Mental Dis. 1985;173:212–20.

33. **Linzer M, Spitzer R, Kroenke K, et al.** Gender, quality of life, and mental disorders in primary care: results from the PRIME-MD 1000 study. Am J Med. 1996;101:526–33.

34. **Takeuchi DT, Chung RC, Lin KM, et al.** Lifetime and twelve-month prevalence rates of major depressive episodes and dysthymia among Chinese Americans in Los Angeles. Am J Psychiatry. 1998;155:1407–14.

35. **Weissman MM, Bland RC, Canino GJ, et al.** Cross-national epidemiology of major depression and bipolar disorder. JAMA. 1996;276:293–9.

36. **Ballenger JC, Davidson JR, Lecrubier Y, et al.** Consensus statement on transcultural issues in depression and anxiety from the International Consensus Group on Depression and Anxiety. J Clin Psychiatry. 2001;13(suppl 62):47–55.

37. **Kleinman A, Kleinman J**. Somatization: the interconnections in Chinese society among culture, depressive experiences, and the meanings of pain In: Kleinman A, Good B, eds. Culture and Depression: Studies in the Anthropology and Cross-Cultural Psychiatry of Affect and Disorder. Berkeley: University of California Press; 1985.

38. **Yen S, Robins CJ, Lin N.** A cross-cultural comparison of depressive symptom manifestation: china and the United States. J Consult Clin Psychol. 2000;68:993–9.

39. **Huang FY, Chung H, Kroenke K, et al.** Racial/ethnic differences in the relationship between depression severity and functional status using the PHQ-9. Psychiatric Services. 2006 (In press).