

Detecting Depression in Older Veterans: Impact of Cultural Diversity on Test Validity

Terry R. Barclay, PhD, Stacy S. Wilkins, PhD, Nancy D. Harada, PhD, Sheryl Osato, PhD, and Shawkat Dhanani, MD

While short forms of the widely used Geriatric Depression Scale have demonstrated usefulness in predominantly white populations, concern has been raised about their accuracy in other cultural groups. This study investigates the issue in a sample of culturally diverse, older veterans.

Depression is a common and disabling disorder among older Americans. The detection and treatment of depressive symptoms in this population is critical to improving quality of life, general health, and morbidity and mortality.^{1,2} Screening instruments for depression are an important component in identifying older adults most likely to have a depressive disorder, and measures are needed that are sensitive and specific for elders of varying cultural backgrounds.

One of the most commonly used depression screening instruments is the Geriatric Depression Scale (GDS).³ Designed over two decades ago, the

original measure contains 30 items answered in a simple yes/no format. While the 30-item GDS is still widely used, several shorter forms have been developed in recent years in an attempt to improve efficiency. Of these, the 15-item and 5-item versions are used most frequently.^{4,5}

The 30-, 15-, and 5-item forms of the GDS all were developed with predominantly white populations, using relatively small sample sizes. The original GDS was based on research involving 47 participants and was later validated on 60 depressed and 40 nondepressed elderly individuals.³ The 15-item form developed by Sheikh and Yesavage was validated on only 35 participants, 17 of whom were in treatment for major depression or a dysthymic disorder.⁴ Demographic information for these two studies is not well reported and indicates only that both men and women were included and that all participants were over the age of 55. The more recent 5-item GDS was developed on 74 male outpatients, the majority of whom were white.⁵

While several studies have shown these instruments to be valid in samples of white individuals,⁶⁻⁸ concerns have been raised about the diagnostic accuracy of the GDS—especially its short forms—in other

cultural groups, including Asian Americans, Mexican Americans, African Americans, and African Caribbean individuals living in the United Kingdom.⁹⁻¹⁴ Various items have been questioned in terms of cultural sensitivity,¹³ ability to distinguish between depressed and nondepressed individuals in certain populations,¹² and comprehensibility for individuals with different educational levels and cultural backgrounds.¹¹

Despite these criticisms, and the lack of data supporting the use of these measures with culturally diverse groups, short forms of the GDS continue to be used routinely to screen for depression in a wide variety of patient populations. The aim of the present study, therefore, was to examine the relationship between cultural status and level of agreement between long and short forms of the GDS.

METHODS

Research participants

A total of 184 community dwelling adults, aged 55 and older, participated in the present study. Participants were enrolled as part of their involvement in the Screening Health Assessment and Preventative Education (SHAPE) program, an outreach service designed to attract older, minority vet-

Dr. Barclay is a postdoctoral research fellow, **Dr. Wilkins** is a staff psychologist for the Geriatric Research, Education and Clinical Center, **Dr. Harada** is a researcher with the office of the associate chief of staff for education, **Dr. Osato** is the director of the Geriatric Psychiatry Outpatient Clinic, and **Dr. Dhanani** is the director of the Geriatric Evaluation and Management Unit, all at the VA Greater Los Angeles Healthcare System, Los Angeles, CA. In addition, Dr. Barclay is a neuropsychology fellow in the department of psychiatry and biobehavioral sciences, Dr. Wilkins is an associate clinical professor of geriatric medicine at David Geffen School of Medicine, Dr. Harada is a professor at the School of Public Health and David Geffen School of Medicine, Dr. Osato is a clinical professor in the department of psychiatry and biobehavioral sciences, and Dr. Dhanani is a clinical professor of geriatric medicine at the David Geffen School of Medicine, all at the University of California, Los Angeles.

erans to the VA health care system. Specific aims of the SHAPE program were to identify individuals who were not utilizing VA services, screen them for selected geriatric conditions, and provide targeted health education. Individuals were recruited for participation in the program through a variety of organizations, including city and county Area Agencies on Aging, veterans' service organizations, and senior centers. Additionally, flyers were posted throughout the community.

Procedure

The Senior SHAPE program was offered both on VA grounds and off-campus in order to widen recruitment and minimize travel requirements for participants with low incomes. All program participants completed interdisciplinary screening by a team of health care professionals trained in geriatrics, including a physician, pharmacist, psychologist, and social

worker. These professionals collected information on demographics, medical history, and functional status and then conducted cognitive and depression screening using the Folstein Mini-Mental State Exam (MMSE) and the 30-item GDS, respectively.^{3,15} Information regarding activity level, number of recent falls, alcohol and tobacco use, and polypharmacy also was obtained. Participants rated their levels of activity on a 6-point Likert scale ranging from 1 (not at all active) to 6 (extremely active). Polypharmacy was defined as the concurrent use of five or more medications.

Results of the screening were examined by a physician and reviewed with each participant by telephone and in writing. Individuals who screened positive for geriatric conditions were provided with disease-specific, printed educational materials and referred for further evaluation and follow-up, as necessary.

Data analyses

For the present study, program data were analyzed using the statistical package SPSS version 11.0 (SPSS, Inc., Chicago, IL). Permission to conduct a retrospective analysis of data for this study was granted by the Institutional Review Board of the West Los Angeles VA Healthcare Center.

Participants who scored below 20 on the MMSE were excluded from data analyses in an effort to avoid inclusion of cognitively impaired participants. One-way analysis of variance (ANOVA) and chi-square tests were used to compare demographic characteristics between the three cultural groups represented in the sample (Asian Americans, African Americans, and whites). Participants were classified as "depressed" or "not depressed" based on normative criteria established for the 30-item GDS using the traditional cutoff score of 11 or higher.³ Scores for the 15- and 5-item

Table 1. Sample demographics and clinical characteristics, by cultural group

Characteristic	African American (n = 42)	Asian American (n = 120)	White (n = 22)	P value
Mean (SD) age (years)	62.5 (8.1)	72.8 (4.8)	73.4 (5.3)	.01 ^a
Mean (SD) level of education (years)	13.2 (2.0)	14.2 (2.7)	15.4 (2.3)	.01 ^a
Mean (SD) activity level ^b	3.3 (1.4)	3.1 (1.0)	3.4 (1.0)	.36
Gender (%)				—
Male	100.0	100.0	100.0	
Living arrangement (%)				.29
With others	82.9	89.1	85.7	
Alone	17.1	10.9	14.3	
Primary source of care (%)				.18
Self	90.5	97.5	95.2	
Others	9.5	2.5	4.8	
Polypharmacy (%)				.37
Yes	2.4	6.7	9.0	
No	97.6	93.3	91.0	

^aStatistically significant at the .01 level. ^bActivity level was rated by participants on a 6-point Likert scale, ranging from 1 (not at all active) to 6 (extremely active).

GDS short forms were obtained by extrapolating data from the full 30-item form, and cutoff values of 5 or higher and 2 or higher, respectively, were used as indicators of depression.

Data were analyzed first using binary logistic regression to examine the relationship between short and long forms of the GDS after controlling for such variables as age and level of education. Diagnostic accuracy, sensitivity, and specificity rates and positive and negative predictive values were then computed for the short forms in each cultural group using 2x2 contingency tables.

The kappa statistic was used to test the level of agreement between the gold standard measure of depression (the 30-item GDS) and the two shorter versions. Kappa values indicate whether the level of agreement between measures is poor (less than 0.2), fair (0.21 to 0.4), moderate (0.41 to 0.6), good (0.61 to 0.8), or very good (0.81 to 1). Internal consistency (reliability) of the 15- and 5-item forms was measured using Cronbach alpha coefficient. Ninety-five percent confidence intervals (CI) were calculated for alpha (CI_{α}) and kappa (CI_{κ}) values. Kappa interval estimation was performed using the approach suggested by Blackman and Koval.¹⁶

RESULTS

Sample demographics

A total of 186 Senior SHAPE program participants completed interdisciplinary screening over an 11-month period. Two participants had MMSE scores below 20 and, therefore, were excluded from further analyses. The remaining 184 participants were selected as the final sample.

Of this sample, 65% were Asian American (primarily second-generation Japanese American), 23% were

African American, and 12% were white (Table 1). All study participants were male. Mean (SD) age of the entire study sample was 70.5 (7.1) years (range, 55 to 86 years). The majority of participants had completed high school, with a mean (SD) of 14.1 (2.6) years of education. Most participants (81%) were living with a spouse or family member at the time of participation, and a large majority (92%) were responsible for their own care. Polypharmacy was reported by 7.6% of participants. Average level of activity was moderate, with a mean (SD) score of 3.13 (1.1) on the 6-point Likert scale.

Using ANOVA and chi-square tests, no differences were found between groups with respect to activity level, living arrangement, primary source of care, or polypharmacy. Pairwise comparisons, however, revealed that African American cultural group was associated with younger age ($F_{2,181} = 52.34, P < .01$) and fewer years of education ($F_{2,175} = 4.94, P < .01$).

Depression classification rates

Total scores obtained by participants on the 30-item GDS ranged from 0 to 21, with an overall mean (SD) score of 3.81 (3.8). Using a cutoff score of 11 or higher, 9.8% of the entire sample was classified as depressed. Broken down by cultural group, depression was identified in 5% of African American participants, 10.8% of Asian American participants, and 13.6% of white participants.

Comparing long and short forms across all participants

In the initial binary logistic regression analyses, use of the 15-item GDS as a predictor yielded a significant overall multivariate model [$\chi^2 (5, n = 178) = 50.4, P < .01$] and correctly classified 95% of cases. Results reveal that, after controlling for age

and education, the 15-item GDS made independent, and statistically significant, contributions to the prediction of depression [$\beta = 4.3, SE = 0.78, Wald = 30.7, P < .01$, odds ratio (OR) = 57.5]. In this model, African American ethnicity was found to be associated with a decreased likelihood of depression [$\beta = -3.7, SE = 1.7, Wald = 4.8, P = .03, OR = 0.02$].

When using the 5-item GDS as a predictor, the overall multivariate model was again significant [$\chi^2 (5, n = 178) = 47.7, P < .01$] and correctly classified 93% of cases. Results revealed that, after controlling for age and education, the 5-item GDS made independent, and statistically significant, contributions to the prediction of depression [$\beta = 3.8, SE = 0.67, Wald = 31.9, P < .01, OR = 45.1$]. In this model, there was a trend for African American ethnicity to be associated with a decreased likelihood of depression—although this relationship was not significant at the .05 level [$\beta = -2.8, SE = 1.5, Wald = 3.3, P = .07, OR = 0.06$].

Test characteristics and kappa statistics by cultural group

Reliability estimates for all three forms of the GDS were obtained using Cronbach alpha. Alpha coefficients for the 30-, 15-, and 5-item forms in the sample as a whole were 0.81 ($CI_{\alpha} = 0.77$ to 0.85), 0.71 ($CI_{\alpha} = 0.65$ to 0.77), and 0.4 ($CI_{\alpha} = 0.26$ to 0.53), respectively. Examination of Cronbach alpha coefficients for the 30-, 15-, and 5-item forms within cultural groups revealed values of 0.74 ($CI_{\alpha} = 0.55$ to 0.87), 0.68 ($CI_{\alpha} = 0.44$ to 0.85), and 0.61 ($CI_{\alpha} = 0.25$ to 0.82), respectively, for white participants; 0.78 ($CI_{\alpha} = 0.67$ to 0.86), 0.6 ($CI_{\alpha} = 0.4$ to 0.76), and 0.01 ($CI_{\alpha} = -0.56$ to 0.41), respectively, for African American participants; and 0.83 ($CI_{\alpha} = 0.78$ to 0.87), 0.75

Table 2. Characteristics of the 15- and 5-item GDS^a short forms, by cultural group

Test characteristic	African American (n = 42)	Asian American (n = 120)	White (n = 22)
GDS-15 (cutoff score ≥ 5)			
Sensitivity ^b	50.0	69.2	66.7
Specificity ^c	90.0	98.1	100.0
PPV ^d	20.0	81.8	100.0
NPV ^e	97.3	96.3	95.0
Diagnostic accuracy ^f	88.1	95.0	95.5
GDS-5 (cutoff score ≥ 2)			
Sensitivity	0.0	84.6	66.7
Specificity	92.5	94.4	100.0
PPV	0.0	64.7	100.0
NPV	94.9	98.1	95.0
Diagnostic accuracy	88.1	93.3	95.5

^aGDS = Geriatric Depression Scale. ^bSensitivity refers to accurate identification of individuals who have the condition. ^cSpecificity refers to accurate identification of individuals who do not have the condition. ^dPPV = positive predictive value. PPV refers to the usefulness of the measure in identifying true positives, given the base rate. ^eNPV = negative predictive value. NPV refers to the usefulness of the measure in identifying true negatives, given the base rate. ^fDiagnostic accuracy refers to the overall accuracy of the measure in correctly identifying true positives and true negatives.

($CI_{\alpha} = 0.67$ to 0.81), and 0.46 ($CI_{\alpha} = 0.29$ to 0.6), respectively, for Asian American participants. Evaluation of other test characteristics revealed that both the 15- and 5-item measures yielded particularly poor sensitivity, positive predictive values, and overall diagnostic accuracy among African American participants (Table 2).

Kappa statistics were computed to evaluate overall level of agreement between the 30-item GDS and the two short forms within each cultural group. Results revealed good levels of agreement between the 15- and 30-item forms in white participants ($\kappa = 0.78$, $CI_{\kappa} = 0.61$ to 0.94 , $P < .01$) and Asian American participants ($\kappa = 0.72$, $CI_{\kappa} = 0.62$ to 0.82 , $P < .01$). Agreement between the two measures was only fair, however, among African American participants ($\kappa = 0.23$, $CI_{\kappa} = -0.02$ to 0.49 , $P = .09$). Likewise, the level of agreement between the 5- and 30-item forms was good with white participants ($\kappa = 0.78$, $CI_{\kappa} =$

0.61 to 0.94 , $P < .01$) and Asian American participants ($\kappa = 0.7$, $CI_{\kappa} = 0.59$ to 0.8 , $P < .01$), but poor among African American participants ($\kappa = -0.06$, $CI_{\kappa} = -0.32$ to 0.2 , $P = .69$).

DISCUSSION

GDS validity in the nonwhite population

Notably, results of this investigation suggest that abbreviated versions of the GDS may not be valid indicators of depression in all cultural groups. More specifically, while the 15- and 5-item versions of the measure compared favorably to the full 30-item form before stratifying study participants by ethnicity, the same was not true once the data were examined more closely within each cultural group. The short forms of the GDS appeared adequate for white and Asian American participants, but these forms appeared inadequate for African American participants. Short

forms of the GDS yielded particularly poor sensitivity, positive predictive power, and overall diagnostic accuracy among individuals in this cultural group.

The findings presented here are consistent with the results of several previous investigations that have examined the utility of GDS short forms in other cultural groups. While most studies to date have focused strictly on white samples, a growing number of investigators have reported poor sensitivity and weak concurrent validity in such populations as African Americans, Mexican Americans, and African Caribbeans.⁹⁻¹¹

These findings are not surprising considering that such measures as the 5-item GDS were developed primarily with white participants and, therefore, do not necessarily take into account the full range of socioeconomic, educational, religious, and cultural factors that may be associated with the manifestation or ex-

pression of depressive symptoms in other groups. For example, there is evidence to suggest that elderly African American and Mexican American individuals tend to express depression more often in terms of somatic symptoms than in the affectively oriented terms typical of elderly white individuals.^{10,17,18} Such a difference may contribute to poor detection of depression in these groups with instruments like the GDS, as somatic items largely were avoided in the development of this measure. Others have pointed out the limitations of using screening instruments like the GDS when educational and cultural factors may alter the interpretation or limit comprehension of questions. For example, Flacker and Spiro found that African American participants in their study sample often asked for clarification during screening and demonstrated a lack of understanding on as many as five questions on the 15-item GDS.¹¹

Study limitations

Results presented here must be considered in light of the study's limitations. First, it should be recognized that the gold standard measure of depression in this investigation was the 30-item GDS and not a clinical diagnosis based on interview. By using the GDS score (and not clinical data) as the gold standard, the authors limited the ability of short forms to perform better than the reference scale.

In addition, the validity of this gold standard and participants' depression classification is unclear. Although the 30-item GDS has been shown to be associated with diagnostic criteria for major depression established by the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) in some groups,¹⁹⁻²¹ it could be argued that even the original version of this screening instrument may not be ap-

propriate for some cultural groups. The fact that depression classification rates in the current study were lowest among African American participants may, in fact, reflect the inadequacy of even the 30-item form to properly identify depression in this group.

It also should be noted that African American participants in our study were somewhat younger and less educated than their Asian American and white counterparts. The extent to which younger age and lower educational level contributed to disagreement between long and short forms of the GDS within this group is unknown. It seems unlikely, however, that age and education alone could account for the findings presented here, as the mean age of African American participants remained greater than 60 years and the mean education for African Americans was above the high school level (13.2 years).

Finally, it is important to recognize that the study's sample consisted of community dwelling, male veterans. Therefore, the extent to which results can be generalized to women, non-veterans, or those in institutional or acute hospital settings remains unclear. Additional validation studies are greatly needed to determine the adequacy of these screening tools in culturally diverse populations.

CONCLUSION

With these caveats in mind, the findings from our investigation have several important implications. First, despite their widespread use, it appears that short forms of the GDS may not be ideal depression screening measures for nonwhite populations. Based on our results, clinicians may improve diagnostic accuracy and sensitivity by using the full 30-item GDS with minority patients.

Alternatively, other measures, such as the Patient Health Question-

naire 9 (PHQ-9), may be considered in settings that serve diverse patient populations. The PHQ-9 is a self-administered version of the Primary Care Evaluation of Mental Disorders (PRIME-MD) that scores each of the nine criteria for major depression from the fourth edition, text revision of the DSM from 0 to 3. The PHQ-9 was validated on 6,000 patients in primary care and obstetrics-gynecology clinics, and no differences were found between African American, Hispanic, or white cohorts.²² Furthermore, in a study of poststroke depression, the PHQ-9 was found to be equally effective for African Americans as for whites (the area under the receiver operating curve was 0.96 for both groups).²³

Finally, our study highlights the fact that the GDS and similar measures are screening instruments and, as such, are not diagnostic of depression. This seems particularly true in the case of those cultural groups in which individuals may not experience or express depression in ways similar to white individuals. The final diagnosis of depression, therefore, requires understanding of unique cultural factors and ultimately remains a clinical judgment. ●

Author disclosures

The authors report no actual or potential conflicts of interest with regard to this article.

Disclaimer

The opinions expressed herein are those of the author and do not necessarily reflect those of Federal Practitioner, Quadrant HealthCom Inc., the U.S. government, or any of its agencies. This article may discuss unlabeled or investigational use of certain drugs. Please review complete prescribing information for specific drugs or drug combinations—including indications,

contraindications, warnings, and adverse effects—before administering pharmacologic therapy to patients.

REFERENCES

1. Bruce ML, McAvay GJ, Raue PJ, et al. Major depression in elderly home health care patients. *Am J Psychiatry*. 2002;159(8):1367–1374.
2. Marzari C, Maggi S, Manzato E, et al. Depressive symptoms and development of coronary heart disease events: The Italian longitudinal study on aging. *J Gerontol A Biol Sci Med Sci*. 2005;60(1):85–92.
3. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: A preliminary report. *J Psychiatry Res*. 1982–1983;17(1):37–49.
4. Sheikh JI, Yesavage JA. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. In: Brink TL, ed. *Clinical Gerontology: A Guide to Assessment and Intervention*. New York, NY: Haworth Press; 1986:165–174.
5. Hoyl MT, Alessi CA, Harker JO, et al. Development and testing of a five-item version of the Geriatric Depression Scale. *J Am Geriatr Soc*. 1999;47(7):873–878.
6. Goring H, Baldwin R, Marriott A, Pratt H, Roberts C. Validation of short screening tests for depression and cognitive impairment in older medically ill inpatients. *Int J Geriatr Psychiatry*. 2004;19(5):465–471.
7. Weeks SK, McGann PE, Michaels TK, Penninx BW. Comparing various short-form Geriatric Depression Scales leads to the GDS-5/15. *J Nurs Scholarsh*. 2003;35(2):133–137.
8. Steiner A, Raube K, Stuck AE, et al. Measuring psychosocial aspects of well-being in older community residents: Performance of four short scales. *Gerontologist*. 1996;36(1):54–62.
9. Abas MA, Phillips C, Carter J, Walter J, Banerjee S, Levy R. Culturally sensitive validation of screening questionnaires for depression in older African-Caribbean people living in south London. *Br J Psychiatry*. 1998;173(9):249–254.
10. Baker FM, Espino DV, Robinson BH. Assessing depressive symptoms in African American and Mexican American elders. *Clin Gerontol*. 1993;14(1):15–29.
11. Flacker JM, Spiro L. Does question comprehension limit the utility of the Geriatric Depression Scale in older African Americans? *J Am Geriatr Soc*. 2003;51(10):1511–1512.
12. Kurlowicz L, Outlaw F, Ratcliffe S, et al. An exploratory study of depression among older African American users of an academic outpatient rehabilitation program. *Arch Psychiatr Nurs*. 2005;19(1):3–9.
13. Mui A. Geriatric Depression Scale as a community screening instrument for elderly Chinese immigrants. *Int Psychogeriatr*. 1996;8(3):445–458.
14. Mui AC, Kang SY, Chen LM, Domanski MD. Reliability of the Geriatric Depression Scale for use among elderly Asian immigrants in the USA. *Int Psychogeriatr*. 2003;15(3):253–271.
15. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state.” A practical method for grading the cognitive status of patients for the clinician. *J Psychiatry Res*. 1975;12(3):189–198.
16. Blackman NJ, Koval JJ. Interval estimation for Cohen’s kappa as a measure of agreement. *Stat Med*. 2000;19(5):723–741.
17. Baker FM, Parker DA, Wiley C, Velli SA, Johnson JT. Depressive symptoms in African American medical patients. *Int J Geriatr Psychiatry*. 1995;10(1):9–14.
18. Gallo JJ, Cooper-Patrick L, Lesikar S. Depressive symptoms of whites and African Americans aged 60 years and older. *J Gerontol B Psychol Sci Soc Sci*. 1998;53(5):277–286.
19. Almeida OP, Almeida SA. Short versions of the Geriatric Depression Scale: A study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. *Int J Geriatr Psychiatry*. 1999;14(10):858–865.
20. Burke WJ, Nitcher RL, Roccaforte WH, Wengel SP. A prospective evaluation of the Geriatric Depression Scale in an outpatient geriatric assessment center. *J Am Geriatr Soc*. 1992;40(12):1227–1230.
21. Lyness JM, Noel TK, Cox C, King DA, Conwell Y, Caine ED. Screening for depression in elderly primary care patients. A comparison of the Center for Epidemiologic Studies–Depression Scale and the Geriatric Depression Scale. *Arch Intern Med*. 1997;157(4):449–454.
22. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613.
23. Williams LS, Brizendine EJ, Plue L, et al. Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke*. 2005;36(3):635–638.