

Published in final edited form as:

Int J Geriatr Psychiatry. 2011 September ; 26(9): 952–960. doi:10.1002/gps.2632.

How do the phq-2, the phq-9 perform in aging services clients with cognitive impairment?

Lisa L. Boyle¹, Thomas M. Richardson¹, Hua He², Yinglin Xia², Xin Tu², Malaz Boustani³, and Yeates Conwell¹

¹Department of Psychiatry, University of Rochester School of Medicine and Dentistry, New York, USA

²Department of Biostatistics, University of Rochester School of Medicine and Dentistry, New York, USA

³Indiana University Center for Aging Research and Regenstrief Institute, Inc, Indiana, USA

Abstract

Objective—To examine the performance of the Patient Health Questionnaire-2 (PHQ-2) and the PHQ-9 in detecting current major depressive episode (MDE) in aging services care management clients who screen positive for cognitive impairment (CI).

Methods—Cross-sectional observational study of 236 care management clients ages 60 years recruited from an Upstate NY aging services agency. The test characteristics of the PHQ-2 and PHQ-9 to screen for depression were calculated using the Structured Clinical Interview for DSM-IV (SCID) to identify MDE (gold standard). CI was identified with the Six-Item Screen (SIS).

Results—Participants had a mean age of 77 years, 68% female, 16% non-white, and 26% had less than a high school education. 16% of participants had CI identified by 2 errors on SIS. Of these, 41% had positive PHQ-2 (scores 3), 43% had positive PHQ-9 (scores 10), while 24% met criteria for MDE. In the sample with CI, the PHQ-2, using a cutoff of 3, had sensitivity=0.78, specificity=0.71, and receiver operating characteristic (ROC) area under the curve (AUC)=0.81, compared with 0.79, 0.82, and 0.88, respectively, for those without CI. In the sample with CI, the PHQ-9, using a cutoff of 10, had sensitivity=0.89, specificity=0.71, and AUC=0.85, compared with 0.85, 0.89, and 0.91, respectively, for those without CI.

Conclusions—Cognitive status should be considered when using the PHQ as a depression screener due to poorer specificity in seniors with CI.

Keywords

depression; cognitive impairment; screening; aging services; care management

Introduction

Comorbid depression and cognitive impairment (CI) are common in older adults and result in loss of quality of life and functioning. 20–32% of community-residing individuals with mild cognitive impairment (MCI) or dementia have comorbid depression (Lyketsos *et al.*, 2000; Lyketsos *et al.*, 2002). Comorbid depression has been associated with increased risk

of MCI's progression to dementia (Modrego and Ferrández, 2004), behavioral problems (Prado-Jean *et al.*, 2010), and nursing home placement (Steele *et al.*, 1990).

To mitigate this public health burden, early detection and treatment of older adults who are at risk for dementia and depression are critical; however, both dementia and depression currently are under-diagnosed and under-treated in primary care settings (Callahan *et al.*, 1994; Callahan *et al.*, 1995; Valcour *et al.*, 2000; Unützer, 2002; Unützer *et al.*, 2002; Boustani *et al.*, 2005), where most community-residing older adults receive their care. Enhancing case identification processes coupled with delivering collaborative care models for seniors with either dementia (Callahan *et al.*, 2006; Vickrey *et al.*, 2006) or depression (Unützer *et al.*, 2002; Hunkeler *et al.*, 2006) improve quality of care and health outcomes. Although dementia collaborative care approaches have been shown to reduce behavioral and psychological symptoms of dementia in general, benefits for comorbid depression have not been demonstrated (Callahan *et al.*, 2006).

Several case recognition screening measures exist for both dementia and depression, yet using these tools together remains a challenge as depression screeners may not perform as well with CI. Identification of brief and effective depression screeners that also are accurate in individuals with CI would allow for more efficient and timely assessment and management of those at risk.

The Patient Health Questionnaire-9 (PHQ-9) (Kroenke *et al.*, 2001) is a popular depression screener that has been used in primary care populations. The PHQ-9 and the PHQ-2 (Whooley *et al.*, 1997; Kroenke *et al.*, 2003; Löwe *et al.*, 2005), a shortened version of the PHQ-9, have been validated for use in older adults (Ell *et al.*, 2005; Li *et al.*, 2007; Lamers *et al.*, 2008; Shah *et al.*, 2009; Watson *et al.*, 2009). The PHQ's performance has been studied in various samples, including the general population (Martin *et al.*, 2006), both older adult home health care clients (Ell *et al.*, 2005) and community-based social services agency clients receiving care management (Richardson *et al.*, in press), and patients with various chronic conditions, including heart (McManus *et al.*, 2005; Stafford *et al.*, 2007; Thombs *et al.*, 2008), liver (Dbouk *et al.*, 2008) and renal (Watnick *et al.*, 2005) diseases, stroke (Williams *et al.*, 2005), and traumatic brain injury (Fann *et al.*, 2005). However, with the exception of a study examining the performance of the PHQ-2 in residential care/assisted living settings (Watson *et al.*, 2009), these studies have either excluded those with significant CI or have not addressed how CI affects PHQ performance. Therefore, the validity of either the PHQ-2 or the PHQ-9 as depression screening tools in cognitively-impaired individuals has not been established. In practice, many older adults with CI may be screened for depression with the PHQ-2 or PHQ-9, so there is a need to establish whether the PHQ is as useful in identifying depression in subjects with CI as it is in those without.

Study aims

The aims of this study were to examine the criterion validity (i.e., sensitivity, specificity, positive and negative predictive values [PPV and NPV], positive and negative likelihood ratios [+LR and -LR], and area under the receiver-operating characteristic [ROC] curve [AUC]) of both the PHQ-2 and the PHQ-9 in a sample of community-dwelling seniors who received in-home assessments for social work care management services, comparing those subjects with and without CI identified with a brief cognitive screener. *Post hoc* analyses for subjects with positive depression screen were conducted to elucidate the depression profiles of both cognitive groups.

Methods

Recruitment of study participants

Subjects were recruited between September 2005 and August 2007 as part of a community-academic research partnership between a regional aging services provider, Eldersource (<http://www.eldersource.org/>), and the University of Rochester Medical Center. Certified by the Council on Accreditation (<http://www.coanet.org/front3/front.cfm?view=7>), Eldersource provides a range of non-medical services, including care management, that help seniors and their families achieve or maintain optimal functioning and remain independent in their homes as long as possible.

Eldersource care managers briefly introduced the study to their clients during the initial home assessments. Clients who consented to be contacted were then screened by telephone for eligibility by the study coordinators. Eligible participants had to be at least 60 years of age and speak English. Participants gave written informed consent at the time of the in-person interview. For subjects who could not give informed consent, proxy consent was obtained. With few exceptions, interviews were conducted in the participants' homes. The Research Subjects Review Board at the University of Rochester approved the study's protocol.

Depression screen

The PHQ-9 (Kroenke *et al.*, 2001) was administered prior to administration of structured diagnostic instruments using the time frame of the prior 2 weeks to assess for depressive symptoms. The PHQ-2 (range 0–6) score was determined by the response to the first two items of the PHQ-9.

Depression gold standard

The Structured Clinical Interview for DSM-IV Psychiatric Disorders (SCID) (First *et al.*, 2002) was used to inform the presence or absence of current major depressive episodes (MDEs). The SCID has been used previously to identify major depression in older adults with dementia (Starkstein *et al.*, 2005a; Teng *et al.*, 2008) and was conducted at the end of the interview. MDE diagnoses were determined by consensus of the interviewers and a geriatric psychiatrist (YC). Given limited access to clients' medical records or other objective measures of health status, we did not attempt to distinguish MDE from depression secondary to medical conditions or substances (Koenig *et al.*, 1997).

Cognitive impairment

The Six-Item Screen (SIS) was used to identify subjects with CI (Callahan *et al.*, 2002). The SIS has been applied as a brief cognitive screener in older adult primary care populations (Boustani *et al.*, 2005; Callahan *et al.*, 2006), in the emergency department (Wilber *et al.*, 2008), and in pre-hospital emergency medical services (Shah *et al.*, 2009). It assesses temporal orientation (three-items) and delayed recall (three-items), an approach similar to one used to screen for CI in a study of seniors receiving a community-based depression intervention (Ciechanowski *et al.*, 2004). The SIS has a range of scores from 0 to 6 errors. A score of ≥ 2 errors was adopted to indicate CI, as this cutoff represented the most balanced tradeoff between sensitivity and specificity in a community-based sample (Callahan *et al.*, 2002).

Statistical analyses

Psychometric properties (i.e., sensitivity, specificity, PPV, NPV, +LR, and –LR) of both the PHQ-2 and the PHQ-9 were calculated for the two cognitive groups (≥ 2 SIS errors vs. <2

SIS errors). In this context, the sensitivity informs the likelihood of a positive PHQ in individuals with MDE; the specificity informs the likelihood of a negative screen if one does not have MDE. The PPV is an estimate of the proportion of individuals who screen positive with the PHQ that are correctly diagnosed as MDE with the SCID, while the NPV is the proportion of individuals with negative PHQ who are correctly identified by the SCID as not having an MDE. Whereas the PPV and NPV are dependent on the sample's prevalence of MDE, the LR+ and LR- are not, and so should be more consistent across sites. LR+ indicates the relative likelihood that a positive screen would be seen in someone with rather than without MDE, while the LR- indicates the relative likelihood that a negative screen would be seen in someone without rather than with MDE. For example, an LR+ of 5–10 would provide moderate increases in post-test probability of depression, as individuals with MDE would be 5–10 times more likely to have a positive screen than those without MDE. In contrast, a screening test with an LR+ of 1 would not be able to discriminate between those with and without disease as those with disease are just as likely to have a positive screen as those without disease. (Grimes and Schulz, 2005). ROC analyses, which provide an overall summary index of the discriminant ability of screening measures, were also conducted to examine AUC, a summary for the ability of the scales to discriminate between depressed and non-depressed subjects, for each of the depression screening methods.

To compare demographic variables that are continuous between the two cognitive groups, either the simple *t*-test for normally distributed variables or the non-parametric Wilcoxon rank test were utilized. To compare categorical demographic variables between the two cognitive groups, the Chi-Square test (or Fisher exact test) was used.

Post hoc descriptive analyses of the means of each of the PHQ-2 and the PHQ-9 items were conducted for positive PHQ-2 and PHQ-9 scores of for 3 and 10, respectively, for each of the two cognitive groups, in order to determine whether differences in distribution of endorsed depressive symptoms may account for observed differences in psychometric performance of the PHQ-9 between groups. These cutoff scores are consistent with clinical convention (Kroenke *et al.*, 2001; Kroenke *et al.*, 2003), and we have previously demonstrated adequate test characteristics using these cutoffs in seniors receiving care management services (Richardson *et al.*, in press). All analyses were performed by using SAS 9.2 (SAS Institute Inc., Cary, NC) and SPSS Statistics 17 (SPSS Inc., Chicago, IL).

Results

During the study period, 1090 clients received in-home care management assessments, of which 643 (59.0%) clients were referred by care managers to study personnel. Some clients were not referred by their care managers to the study because they had no further interest in Eldersource services; others were not referred due to fluctuating size of caseloads and staff turnover (higher caseloads were associated with lag in study referral rates). Of the 643 clients referred, study staff did not contact 63 clients due to very high referral rates exceeding study resources. Study personnel could not reach 47 clients, and 24 had moved, died or were ineligible based on age or language criteria. Of the remaining 509 eligible subjects, 131 (25.7%) declined participation. 378 clients provided informed consent and were enrolled in the study. Examination of Eldersource administrative data revealed no statistically significant differences in age, gender, income, marital status, and race between clients who enrolled and those who did not. Of these 378 subjects, the first 142 enrolled were not administered the SIS and were excluded from these analyses. Comparisons between the 236 with and the 142 without SIS revealed no statistically significant differences in age, gender, income, marital status, race, PHQ-2 and PHQ-9 scores, or frequency of MDE; however, those with SIS were more likely to have higher levels of

education (26.3% vs. 35.9% with <12 y education for those with vs. without SIS, respectively; $\chi^2=3.924$, $df=1$, $p=0.048$).

As shown in Table 1, the majority of subjects were female, white, relatively well-educated, and with household incomes less than \$2000/month. Other than the CI group being older than those without CI, there were no other statistically significant differences in the demographic variables shown in Table 1. The SIS, PHQ-2, PHQ-9, and SCID results are summarized in Table 2.

PHQ-2 and PHQ-9

The psychometric properties for the PHQ-2 and the PHQ-9 within the two cognitive groups are reported in Table 3. For both groups the conventional cutoff scores of 3 for the PHQ-2 and 10 for the PHQ-9 appear to provide an optimal balance of sensitivity and specificity. Relative to the cognitively intact group, there were drops in specificity and PPV for both the PHQ-2 and the PHQ-9 in those with CI. However, the difference in psychometric performance between subjects with and without CI was less pronounced using the PHQ-2 than the PHQ-9. For both the PHQ-2 (Table 3a) and PHQ-9 (Table 3b), the difference in LR + was substantial in both groups.

Figure 1 illustrates the ROC curves and AUC for the PHQ-2 and PHQ-9 in the participants (a) with and (b) without CI. The curves demonstrate that both the PHQ-2 and PHQ-9 had poorer overall performance in participants with CI than those without CI.

In order to examine whether specific PHQ-9 items accounted for the differences observed in the scale's psychometric properties in the two cognitive groups, we conducted a comparison between cognitive groups of items that were endorsed in individuals who scored PHQ-9 10. As this particular analysis was designed *post hoc*, we did not conduct statistical comparisons. Marked differences existed among participants who scored 10 on the PHQ-9 between the two cognitive groups in four specific items of the measure. Changes in appetite (mean [\pm SD]=1.61 \pm 1.13 for non-CI group; mean=1.00 \pm 1.15 for CI group), feeling bad about oneself or as if one is a failure (mean=1.56 \pm 1.27 for non-CI group; mean=0.88 \pm 1.09 for CI group), and thoughts of death or self-harm (mean=0.63 \pm 0.98 for non-CI group; mean=0.31 \pm 0.48 for CI group) were endorsed more often in the non-CI subjects, while concentration problems were more often endorsed by subjects with CI (mean=1.23 \pm 1.25 for CI group; mean=0.200 \pm 1.32 for non-CI group). Similar comparison of the two items of the PHQ-2 in those who scored 3 revealed no appreciable differences in response between the two cognitive groups (loss of interest: mean 2.06 \pm 0.93 for non-CI and mean=2.33 \pm 0.72 for CI; down mood: mean=2.12 \pm 0.93 for non-CI and mean=1.87 \pm 1.23 for CI group).

Discussion

To our knowledge, this study is the first to examine the criterion validity of the PHQ-2 and PHQ-9 among aging services clients receiving in-home social work assessments who have CI as identified by a brief cognitive screener. We found that there were differences in performance for both the PHQ-2 and the PHQ-9 in clients with CI compared with those without CI. The overall performances of the PHQ-2 and PHQ-9 as characterized by their AUC was good (>0.8) in both subgroups of the sample.

At the same time, however, there were differences between the two measures that may inform decisions about their use in this setting. The sensitivity of the PHQ-2 (0.78–0.79) was lower than the PHQ-9 (0.85–0.89) for both cognitive groups. However, the specificity of the PHQ-9 was 18% lower in subjects with CI than in those without CI (0.71 vs. 0.89)

while the decrement in specificity of the PHQ-2 when used in clients with (0.74) and without CI (0.82) was less pronounced. In general, the PHQ-2 and PHQ-9 using conventional cut-offs performed better in the non-CI group based on a sharper drop in LR+ in the CI group. Using conventional cut-offs, the PHQ-2 and the PHQ-9 perform poorly in the CI group based on the LR+ of <5 (Grimes and Schulz, 2005). The magnitude of drop in LR+ between the two cognitive groups was larger with the PHQ-9 compared with the PHQ-2 which may further reflect limitations of the longer scale when applied to the CI group. The cut-off score would have to be increased to 14 for the PHQ-9, to achieve a moderate increase in post-test probability LR+ of 5–10 (Grimes and Schulz, 2005).

Post hoc analyses in those with positive PHQ screens of PHQ-9 items endorsed by CI and non-CI groups revealed disparate scores for the concentration, feelings of failure, appetite change, and thoughts of death or self-harm items. CI subjects endorsed higher levels of difficulties with concentration or thinking, as one would expect, and lower levels of perceived failure and thoughts of death or self-harm. The difference in appetite change is more difficult to explain. Nonetheless, the fact that none of these three items is included in the PHQ-2 may explain why there was a less pronounced drop-off in specificity when using the scale with CI subjects compared with the PHQ-9.

Interpretation of our findings should be made in light of the following limitations. Those who completed the cognitive screener had higher levels of education compared with the first 142 participants who were not screened for CI and therefore were not included in the analyses. The group who screened positive for CI was older than the group without positive cognitive screen. Our sample had a relatively small number of individuals with identified CI and even smaller numbers of those with both CI and MDE. The lower prevalence of CI in the sample may reflect selection bias, as seniors who are most cognitively-impaired, depressed, disabled, or frail may not have been recruited into our study. Our findings may not generalize to other populations, as we had a mostly white and relatively well-educated sample derived solely from one community-based aging services provider. Our study did not administer the SCID to caregiver informants which may lead to under-recognition of depression in the CI group. Our use of a brief cognitive screening tool as a proxy measure for CI instead of more extensive neuropsychometric and clinical evaluations may have resulted in less precise classification of our participants' cognitive status.

However, we argue that our findings contribute to our understanding of how to interpret the results of the PHQ-2/-9 when co-administered with a CI screening instrument. As both CI and depression are prevalent in older adults, future comprehensive care models will need to concurrently assess for a variety of common neuropsychiatric conditions.

Several challenges exist for clinicians caring for seniors with depression and CI. Whether to adopt widespread screening remains controversial. The U.S. Preventative Services Task Force (USPSTF) has recommended depression screening of adults when resources are available to allow for accurate diagnosis and adequate depression treatment and follow-up (USPSTF, 2009), while they have not recommended routine dementia screening in primary care due to insufficient evidence to determine whether benefits of screening outweigh potential harms (USPSTF, 2003). Moreover, the lack of consensus about diagnostic criteria for depression in individuals with Alzheimer's Disease (AD) poses another barrier for clinicians. Provisional criteria for depression of AD have been proposed as an alternative to the DSM-IV (Olin *et al.*, 2002; Rosenberg *et al.*, 2005; Teng *et al.*, 2008) yet have not been adopted widely. Additionally, discriminating between apathy and depression in individuals with dementia also poses challenges for clinicians as these can be distinct entities but also may overlap (Starkstein *et al.*, 2005b). Furthermore, demonstrated effectiveness of antidepressant treatments for depression associated with AD has not been established.

Recently, the selective serotonin reuptake inhibitor, sertraline, was not found to be efficacious in treating depression of AD (Rosenberg *et al.*, 2010).

Turning to the challenge of assessment of depression in individuals with CI or dementia, there are a variety of depression screeners that have been used in older adults; however, the validity of using brief depression screening tools in older adults with dementia has not been as well established. The Cornell Scale for Depression in Dementia is an example of a clinician-administered instrument specifically designed to assess for depression in those with dementia (Alexopoulos *et al.*, 1988); however, it is less practical as a screening instrument given the longer administration time. A version of the Cornell Scale for Depression in Dementia adapted for direct care providers in residential care/assisted living settings performed poorly (Watson *et al.*, 2009). The Geriatric Depression Scale (GDS) is another well-validated depression screener that was specifically designed for use in older adults. It has abbreviated forms (5- and 15-items) that make it attractive as a screening tool (Rinaldi *et al.*, 2003; Sheikh and Yesavage, 1986). However, the GDS demonstrated poor validity when used in a population of patients with mild AD (Burke *et al.*, 1989). The GDS had favorable test characteristics in a group of older stroke patients with no or only mild to moderate CI; however, accuracy decreased in those with more CI (i.e., lower Mini-Mental State Examination scores) (Agrell and Dehlin, 1989).

Our findings suggest that the PHQ-2 and PHQ-9 have limitations as depression screeners for individuals with CI as well. However, these screeners could be a useful measure for first-stage screening for depression as they perform adequately in clients with and without CI (ROC AUC > 0.8) in this sample with approximately 24% prevalence of depression. However, when determining the performance of the PHQ-2 and PHQ-9 using the LR+, which is independent of a sample's prevalence of depression, to achieve moderate increases in the post-test probability of depression one would have to increase the cut-off scores for both of these instruments in those with CI. If these cut-offs were applied, there would be compromise of the sensitivity of the screen in those with CI but improved specificity that would result in a decrease in the frequency of false-positives. The PHQ's relatively low specificity in seniors with CI at conventional cut-offs indicates that both the PHQ-2 and the PHQ-9 should either be used with a second-stage screening measure that offers higher specificity or followed by a clinical evaluation. Moreover, compared with the PHQ-9, the PHQ-2's brevity and ease of administration suit it well to busy social service settings and will support its uptake and dissemination, provided that there is a system in place for second-stage screening and assessment (e.g., referral to the primary care provider).

Conclusion

The cognitive status of individuals should be considered when interpreting the results of the PHQ depression screen because it does not perform as well in seniors with CI using the usual cutoff scores. Although the performance of the PHQ-2 and the PHQ-9 are not optimal in those with CI, other depression screeners that are brief and valid in individuals with significant CI have not been identified. Therefore, adjustments should be made for the PHQ-2 and the PHQ-9 by either applying a higher cut-off score for those with CI or if choosing to use conventional cut-off scores, then pairing the PHQ with a second-line depression scale (such as the Cornell Scale for Depression in Dementia) or referral for clinical assessment. For agencies with resources available that can absorb the costs associated with false-positive screens, the PHQ-9 could be used to maximize sensitivity to detect depression in those with CI. Future work to develop a consensus for depression screening tools to be used for individuals with CI that balances need for brevity and validity is needed.

Acknowledgments

This research was supported in part by grants from NIMH (T32 MH073452 and R24MH071604), AHRQ (T32 HS000044-15) and the American Foundation for Suicide Prevention. The authors thank the entire Eldersource staff for their contributions and for making this work possible, and Constance Bowen and Judy Woodhams for their invaluable assistance in data collection. This work was presented in part at the American Association for Geriatric Psychiatry, Honolulu, HI, March 7, 2009.

References

- Agrell B, Dehlin O. Comparison of six depression rating scales in geriatric stroke patients. *Stroke*. 1989; 20:1190–1194. [PubMed: 2772980]
- Alexopoulos GS, Abrams RC, Young RC, et al. Cornell scale for depression in dementia. *Biol Psychiatry*. 1988; 23:271–284. [PubMed: 3337862]
- Boustani M, Callahan CM, Unverzagt FW, et al. Implementing a screening and diagnosis program for dementia in primary care. *J Gen Intern Med*. 2005; 20:572–577. [PubMed: 16050849]
- Burke WJ, Houston MJ, Boust SJ, et al. Use of the geriatric depression scale in dementia of the Alzheimer type. *JAGS*. 1989; 37:856–860.
- Callahan CM, Hendrie HC, Dittus RS, et al. Improving treatment of late life depression in primary care: a randomized clinical trial. *JAGS*. 1994; 42:839–846.
- Callahan CM, Hendrie HC, Tierney WM. Documentation and evaluation of cognitive impairment in elderly primary care patients. *Ann Intern Med*. 1995; 122:422–429. [PubMed: 7856990]
- Callahan CM, Unverzagt FW, Hui SL, et al. Six-item screener to identify cognitive impairment among potential subjects for clinical research. *Med Care*. 2002; 40:771–781. [PubMed: 12218768]
- Callahan CM, Boustani MA, Unverzagt FW, et al. Effectiveness of collaborative care for older adults with Alzheimer disease in primary care: a randomized controlled trial. *JAMA*. 2006; 295:2148–2157. [PubMed: 16684985]
- Ciechanowski P, Wagner E, Schmaling K, et al. Community-integrated home-based depression treatment in older adults: a randomized controlled trial. *JAMA*. 2004; 291:1569–1577. [PubMed: 15069044]
- Dbouk N, Arguedas MR, Sheikh A. Assessment of the PHQ-9 as a screening tool for depression in patients with chronic hepatitis C. *Dig Dis Sci*. 2008; 53:1100–1106. [PubMed: 17934817]
- Ell K, Unutzer J, Aranda M, et al. Routine PHQ-9 depression screening in home health care: depression, prevalence, clinical and treatment characteristics and screening implementation. *Home Health Care Serv Q*. 2005; 24:1–19. [PubMed: 16446263]
- Fann JR, Bombardier CH, Dikmen S, et al. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. *J Head Trauma Rehabil*. 2005; 20:501–511. [PubMed: 16304487]
- First, MB.; Spitzer, RL.; Gibbon, M. Structured Clinical Interview for DSM-IV-TR Axis I Disorders. Research Version, Non-patient Edition. Biometrics Research, New York State Psychiatric Institute; New York: 2002. (SCID-I/NP)
- Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet*. 2005; 365:1500–1505. [PubMed: 15850636]
- Hunkeler EM, Katon W, Tang L, et al. Long term outcomes from the IMPACT randomised trial for depressed elderly patients in primary care. *BMJ*. 2006; 332:259–263. [PubMed: 16428253]
- Koenig H, George L, Peterson B, et al. Depression in medically ill hospitalized older adults: prevalence, characteristics, and course of symptoms according to six diagnostic schemes. *Am J Psychiatry*. 1997; 154:1376–1383. [PubMed: 9326819]
- Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001; 16:606–613. [PubMed: 11556941]
- Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003; 41:1284–1292. [PubMed: 14583691]

- Lamers F, Jonkers CCM, Bosma H, et al. Summed score of the Patient Health Questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. *J Clin Epidemiol.* 2008; 61:679–687. [PubMed: 18538262]
- Li C, Friedman B, Conwell Y, et al. Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *JAGS.* 2007; 55:596–602.
- Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res.* 2005; 58:163–171. [PubMed: 15820844]
- Lyketsos CG, Lopez O, Jones B, et al. Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: results from the Cardiovascular Health Study. *JAMA.* 2002; 288:1475–1483. [PubMed: 12243634]
- Lyketsos CG, Steinberg M, Tschanz JT, et al. Mental and behavioral disturbances in dementia: findings from the Cache County Study on Memory in Aging. *Am J Psychiatry.* 2000; 157:708–714. [PubMed: 10784462]
- Martin A, Rief W, Klaiberg A, et al. Validity of the brief Patient Health Questionnaire mood scale (PHQ-9) in the general population. *Gen Hosp Psychiatry.* 2006; 28:71–77. [PubMed: 16377369]
- McManus D, Pipkin SS, Whooley MA. Screening for depression in patients with coronary heart disease (data from the Heart and Soul Study). *Am J Cardiol.* 2005; 96:1076–1081. [PubMed: 16214441]
- Modrego PJ, Ferrández J. Depression in patients with mild cognitive impairment increases the risk of developing dementia of Alzheimer type: a prospective cohort study. *Arch Neurol.* 2004; 61:1290–1293. [PubMed: 15313849]
- Olin JT, Katz IR, Meyers BS, et al. Provisional diagnostic criteria for depression of Alzheimer disease: rationale and background. *Am J Geriatr Psychiatry.* 2002; 10:129–141. [PubMed: 11925274]
- Prado-Jean A, Couratier P, Druet-Cabanac M, et al. Specific psychological and behavioral symptoms of depression in patients with dementia. *Int J Geriatr Psychiatry.* 2010 published online (Accessed March 13, 2010).
- Richarson TM, He H, Podgorski C, et al. Screening depression aging services clients. *Am J Geriatr Psychiatry.* in press.
- Rinaldi P, Mecocci P, Benedetti C, et al. Validation of the five-item Geriatric Depression Scale in elderly subjects in three different settings. *JAGS.* 2003; 51:694–698.
- Rosenberg PB, Drye LT, Martin BK, et al. Sertraline for the treatment of depression in Alzheimer disease. *Am J Geriatr Psychiatry.* 2010; 18:136–145.
- Rosenberg PB, Onyike CU, Katz IR, et al. Clinical application of operationalized criteria for ‘depression of Alzheimer disease’. *Int J Geriatr Psychiatry.* 2005; 20:119–127. [PubMed: 15660424]
- Shah MN, Karuza J, Rueckmann E, et al. Reliability and validity of prehospital case finding for depression and cognitive impairment. *JAG.* 2009; S57:697–702.
- Sheikh, JI.; Yesavage, JA. *Clinical Gerontology: A Guide to Assessment and Intervention.* Haworth Press; New York: 1986. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version; p. 165-173.
- Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Geriatric Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. *Gen Hosp Psychiatry.* 2007; 29:417–424. [PubMed: 17888808]
- Starkstein SI, Ingram L, Garau ML, et al. On the overlap between apathy and depression in dementia. *J Neurol Neurosurg Psychiatry.* 2005b; 76:1070–1074. [PubMed: 16024880]
- Starkstein SE, Jorge R, Mizrahi R, et al. The construct of minor and major depression in Alzheimer’s disease. *Am J Psychiatry.* 2005a; 162:2086–2093. [PubMed: 16263848]
- Steele C, Rovner B, Chase GA, et al. Psychiatric symptoms and nursing home placement of patients with Alzheimer’s disease. *Am J Psychiatry.* 1990; 147:1049–1051. [PubMed: 2375439]
- Teng E, Ringman JM, Ross LK, et al. Diagnosing depression in Alzheimer disease with the National Institute of Mental Health provisional criteria. *Am J Geriatr Psychiatry.* 2008; 16:469–477. [PubMed: 18515691]

- Thombs BD, Ziegelstein RC, Whooley MA. Optimizing detection of major depression among patients with coronary artery disease using the Patient Health Questionnaire: data from the Heart and Soul Study. *J Gen Intern Med.* 2008; 23:2014–2017. [PubMed: 18815842]
- Unützer J. Diagnosis and treatment of older adults with depression in primary care. *Biol Psychiatry.* 2002; 52:285–292. [PubMed: 12182933]
- Unützer J, Katon W, Callahan CM, et al. the IMPACT Investigators. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA.* 2002; 288:2836–2845. [PubMed: 12472325]
- U.S Preventive Task Services Force. Screening for dementia: recommendation and rationale. *Ann Intern Med.* 2003; 138:925–926. [PubMed: 12779303]
- U.S Preventive Task Services Force. Screening for depression in adults: US Preventive Services Task Force recommendation statement. *Ann Intern Med.* 2009; 151:784–792. [PubMed: 19949144]
- Valcour VG, Masaki KH, Curb JD, et al. The detection of dementia in the primary care setting. *Arch Intern Med.* 2000; 160:2964–2968. [PubMed: 11041904]
- Vickrey BG, Mittman BS, Connor KI, et al. The effect of a disease management intervention on quality and outcomes of dementia care: a randomized, controlled trial. *Ann Intern Med.* 2006; 145:713–726. [PubMed: 17116916]
- Watnick S, Wang PL, Demadura T, et al. Validation of two depression screening tools in dialysis patients. *Am J Kidney Dis.* 2005; 46:919–924. [PubMed: 16253733]
- Watson LC, Zimmerman S, Cohen LW, et al. Practical depression screening in residential care/assisted living: five methods compared with gold standard diagnoses. *Am J Geriatr Psychiatry.* 2009; 17:556–564. [PubMed: 19554670]
- Whooley MA, Avins AL, Miranda J, et al. Case-finding instruments for depression: two questions are as good as many. *J Gen Intern Med.* 1997; 12:439–445. [PubMed: 9229283]
- Wilber ST, Carpenter CR, Hustey FM. The Six-Item Screener to detect cognitive impairment in older emergency department patients. *Acad Emerg Med.* 2008; 15:613–616. [PubMed: 18691212]
- Williams LS, Brizendine EJ, Plue L, et al. Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke.* 2005; 36:635–638. [PubMed: 15677576]

Key Points

- Cognitive status should be considered when using the PHQ-2 or PHQ-9 as depression screeners due to poorer specificity in seniors with cognitive impairment.
- Future work to develop a consensus for depression screening tools to be used for individuals with cognitive impairment that balances need for brevity and validity is needed.

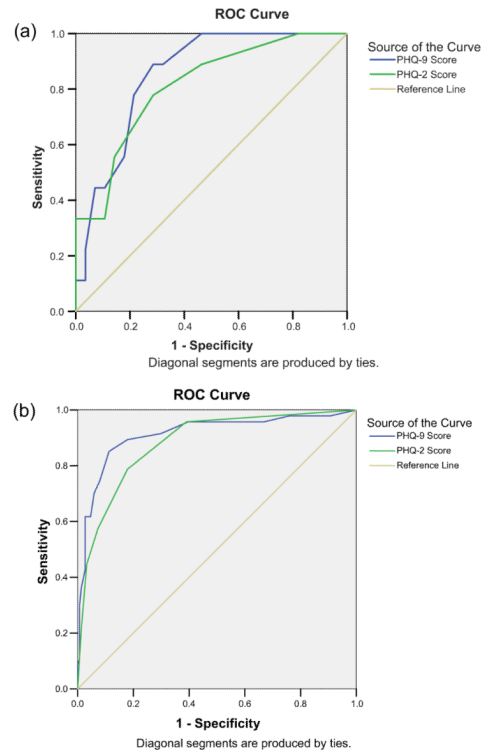


Figure 1. Performance of PHQ-2 and PHQ-9 related to cognitive status. a) SIS \geq 2 errors. b) SIS $<$ 2 errors. Patient Health Questionnaire-2 (PHQ-2); Patient Health Questionnaire-9 (PHQ-9); Six-Item Screen (SIS); receiver operating characteristic (ROC) curve. This figure is available in colour online at wileyonlinelibrary.com/journal/gps

Table 1

Sample demographics

Baseline characteristics	Total (n=236)	SIS 2 errors (n=37)	SIS <2 errors (n=199)	p-value
Age, mean±SD range	76.5±9.2 60–100	81.3±9.4 60–96	75.6±8.9 60–100	0.0010
Female, no. (%)	162 (68)	20(54)	142 (71)	0.0524
Education <12 year, no. (%)	62 (26)	13 (35)	49 (25)	0.2217
Married, no. (%)	97 (41)	18 (49)	79 (40)	0.3639
Non-white, no. (%)	39 (16)	7 (19)	32 (16)	0.6353
Income \$2K/month, no. (%)	170 (73)	25 (69)	145 (74)	0.5464

Note: Six-Item Screen (SIS). *p*-values reported for comparisons of continuous demographic variables between the two cognitive groups utilized either the simple *t*-test for normally distributed variables or the non-parametric Wilcoxon rank test. The Chi-Square test (or Fisher exact test) was used to compare categorical demographic variables between the two cognitive groups.

Table 2

Cognition and depression measures

Baseline characteristics	Total	SIS 2 errors	SIS <2 errors
SIS, mean±SD, range	0.8±1.3, 0–6	3.3±1.5, 2–6	0.3±0.5, 0–1
PHQ-9, mean±SD, range	7.7±5.4, 0–27	8.6±5.1, 0–21	7.5±5.5, 0–27
PHQ-9 10, no. (%)	73 (31)	16 (43)	57 (29)
PHQ-2, mean±SD, range	2.1±1.8, 0–6	2.3±1.8, 0–6	2.0±1.8, 0–6
PHQ-2 3, no. (%)	79 (34)	15 (41)	64 (32)
MDE by SCID criteria, no. (%)	57 (24)	9 (24)	48 (25)

Note: Six-Item Screen (SIS); Patient Health Questionnaire-9 (PHQ-9); major depressive episode (MDE); Structured Clinical Interview for DSM-IV (SCID).

Table 3

Characteristics of the (a) PHQ-2 and (b) PHQ-9 in Subjects with and without CI

(a) PHQ-2						
With CI (SIS ≥ 2 errors)						
PHQ-2 scores	Sensitivity	Specificity	PPV	NPV	LR+	LR-
1	1.00	0.18	0.28	1.00	1.22	0.00
2	0.89	0.54	0.38	0.83	1.92	0.21
3	0.78	0.71	0.46	0.76	2.72	0.31
4	0.56	0.86	0.55	0.66	3.89	0.52
5	0.33	0.89	0.50	0.57	3.11	0.75
6	0.33	1.00	1.00	0.69	–	0.67
Without CI (SIS <2 errors)						
PHQ-2 scores	Sensitivity	Specificity	PPV	NPV	LR+	LR-
1	0.98	0.30	0.31	0.93	1.39	0.07
2	0.96	0.61	0.44	0.93	2.45	0.07
3	0.79	0.82	0.58	0.79	4.40	0.26
4	0.57	0.93	0.72	0.69	7.89	0.46
5	0.45	0.97	0.81	0.64	13.49	0.57
6	0.21	0.99	0.84	0.56	16.06	0.80
(b) PHQ-9						
With CI (SIS ≥ 2 errors)						
PHQ-9 scores	Sensitivity	Specificity	PPV	NPV	LR+	LR-
7	1.00	0.46	0.37	1.00	1.87	0.00
8	1.00	0.54	0.41	1.00	2.16	0.00
9	0.89	0.68	0.47	0.95	2.77	0.16
10	0.89	0.71	0.50	0.95	3.11	0.16
11	0.78	0.79	0.54	0.92	3.64	0.28
12	0.56	0.82	0.50	0.85	3.11	0.54
13	0.44	0.89	0.57	0.62	4.15	0.62
14	0.44	0.93	0.67	0.63	6.25	0.60
Without CI (SIS <2 errors)						
PHQ-9 scores	Sensitivity	Specificity	PPV	NPV	LR+	LR-
7	0.96	0.60	0.43	0.98	2.41	0.07
8	0.91	0.70	0.49	0.96	3.07	0.12
9	0.89	0.82	0.61	0.96	5.00	0.13
10	0.85	0.89	0.71	0.95	7.56	0.17
11	0.74	0.92	0.75	0.92	9.37	0.28
12	0.70	0.94	0.79	0.91	11.78	0.32

13	0.62	0.95	0.81	0.71	13.31	0.40
14	0.62	0.97	0.88	0.72	23.29	0.39

Note: Patient Health Questionnaire-2 (PHQ-2); Patient Health Questionnaire-9 (PHQ-9); cognitive impairment (CI); Six-Item Screen (SIS); positive predictive value (PPV); negative predictive value (NPV); positive likelihood ratio (LR+); negative likelihood ratio (LR-).

The bolded numbers represent the test characteristics using conventional cutoff scores of 3 for the PHQ-2 and 10 for the PHQ-9.