informa
healthcare

SCREENING INSTRUMENTS FOR OLDER ADULT DEPRESSIVE DISORDERS: UPDATING THE EVIDENCE-BASED TOOLBOX

Marian W. Roman, PhD, PMHCNS-BC
Bonnie L. Callen, PhD, PCHCNS-BC
University of Tennessee, College of Nursing, Knoxville, Tennessee, USA

The prevalence of clinically significant depressive disorders in persons 65 years of age or older in the United States has been estimated to be approximately 15%, increasing the risk for functional decline, morbidity, and mortality. Utilization of a well-chosen screening instrument has been shown to improve the rates of recognition of depressive disorders in older adults. This paper presents a targeted review of the most commonly accepted tools for case-finding of depressive disorders in older adults. After a review of the benefits and shortcomings of screening tools, the strengths, weaknesses, and utility of selected depression scales in geriatric clinical settings are discussed.

In 2002 the prevalence of clinically significant depressive disorders in persons 65 years of age or older in the United States was estimated to be approximately 15% (Federal Interagency Forum, 2006). Research indicates that rates for recognition of mood disorders in primary care settings remain poor, and only a small percentage of older adults are seen by specialists in geriatrics or mental health. The prudent use of one of several well tested screening tools may aid in recognizing the likelihood of a depressive disorder in an older adult, particularly when symptoms may be subtle, masked, or confounded by comorbidities.

Ironically, the screening instruments reviewed here are more frequently part of standard practice in specialty, versus primary, care settings. Specialists in geriatrics or geropsychiatry may be more cognizant

Address correspondence to Marian W. Roman, 1200 Volunteer Blvd., Knoxville, TN 17996-4180. E-mail: mroman@utk.edu

of the spectrum of such disorders and their subtle and myriad manifestations than primary care providers; nevertheless, they regularly use instruments when conducting an assessment as well monitoring changes over time.

All the instruments discussed are evidence-based. They have been conceptually validated, tested specifically with older adult populations, have high rates of both sensitivity and specificity (from mid-80% to 100%), and have a high predictability for eventual diagnosis when designated cut-points are used.

The use of these instruments in clinical practice is quite different from their use in research. Clinicians are interested in an individual patient's score and may choose to repeat the measure to ascertain individual change over time. Researchers utilize these scales to measure prevalence of depression within groups, the mean difference in scores between two groups, or pre- and post-intervention.

Psychometricians suggest that a high internal reliability rate (.9–.95) is necessary for a tool to be valid in measuring change within a single individual over time (Kaplan & Sacuzzo, 1989); few tools meet this standard. Indeed, there are no published studies of any of the instruments reviewed herein reporting this degree of internal reliability precision. Nevertheless, some health services researchers recognize that, based on the context of clinical practice and using practice-based wisdom, clinicians may find that flexibility is needed in interpreting and understanding patients' scores in real world situations in which variables are not controlled and contextual factors influence scores (McHorney & Tarlov, 1995). The use of screening tools in the clinical context is the basis for this review.

PROCESS AND CAVEATS

Screening is not the equivalent of diagnosis. Despite what sometimes appears to be the use of screening tool scores as a proxy for diagnosis (in both research and practice settings), screening instruments are employed to facilitate case-finding and the diagnosis of conditions. Their utilization should be directed and chosen by a professional, although most instruments may be self-administered and scored by non-licensed personnel. Instruments assist the practitioner; they do not substitute for diagnostic interviews or other objective diagnostics. Alone or in tandem, no screening tool is diagnostic; it suggests the presence or absence of suspicious indicators as compared to population norms. Comprehensive diagnostic workups may be triggered by the results of such tools and it is these workups that are the basis for a diagnosis.

The clinician's approach and general preliminary impressions of the client form the foundation for any patient encounter and can influence the accuracy of the screening or diagnostic process. Information can be gleaned from listening and observing demeanor, dress, and casual speech. The clinician's ability to establish trust and decrease anxiety in the client cannot be underestimated. Fears of loss of identity and independence are paramount in older adults, and the anxiety surrounding testing may interfere with concentration, persistence, and performance. Reassurance that screening for mood problems is just good practice, similar to routine breast or prostate exams, can aid in anxiety reduction.

A few general principles in the assessment of older adults are worthy of reiteration (Gallo, Fulmer, Paveza, & Reichel, 2000; Lichtenberg, 1999). Concerning the older individual, three words to remember are: heterogeneity, context, and functioning. As we age, we become more dissimilar. Our inter-individual differences increase as a result of the cumulative effects of environment and experiences over time. Life experiences, distal and proximal, must be considered when assessing older adults. Educational attainment is the most frequently considered contextual influence when norms are set for screening tools; however, other contexts include ethnic identity and previous life situations, including former work and home settings. Contemporaneous individual contexts also must be considered, including the extent of sensory impairments, recent losses, or relocations.

Population and age-based norms are inherent in the use of evidence-based instruments; however, due to heterogeneity, it may be much more salient to compare a patient's current functioning to his or her own previous functional level or previous history of successful adaptation than to a norm. Although the designated cut-points of scales have been shown in clinical trials to maximize sensitivity with acceptable specificity in a population, the difference of one point over or under in a particular individual may seem arbitrary, and additional individualized contextual information should be obtained to complement screening scores.

When choosing among instruments, three words to remember are: redundancy, brevity, and evidence-based (Lichtenberg, 1999). Redundancy refers to the use of several measures to increase the sensitivity and specificity of overall findings. Brevity is important for both the busy patient, who may tire easily, and the busy clinical staff. The instruments reviewed herein meet the brevity and evidence-based criteria and may complement each other in detecting a possible depressive mood syndrome.

Most screening instruments utilized to detect mental health disorders have been validated by a comparison against a comprehensive diagnosis performed by a professional. The Structured Clinical Interview (SCID;

First & Gibbon, 2004), derived from criteria from the *Diagnostic and Statistical Manual for Mental Disorders-IV* (DSM-IV; American Psychiatric Association, 2000), is usually the gold standard comparison. Participants' scores on an instrument are compared to their SCID results for concordance. The operating characteristics of sensitivity and specificity rates for a certain instrument's optimal score are derived from this testing. The sensitivity rate for a screening tool is equal to the percentage of participants with confirmed clinical disorders who had been correctly detected with that tool using a specified cut-point. The specificity rate, a smaller percentage, refers to the rate of detected cases that specifically met the criteria for the chosen diagnostic category, and not another disorder, such as a primary anxiety or dementia. To err on the side of false positives (sensitivity) is preferable to being too specific when the purpose is screening (Lichtenberg, 1999).

TOOLS FOR THE TOOLBOX: SCREENING FOR DEPRESSIVE DISORDERS

Clinically significant depression is a spectrum disorder rather than a categorical disease entity. Subsyndromal depression has a variety of names (e.g., subsyndromal symptomatic depression, subclinical depression, mild depression, subdysthymic depression) and has been estimated to affect up to 10% of older adults (Liebowitz et al., 1997; Lyness et al., 1997), with a 3% prevalence of major depressive disorder in healthy elderly persons living in the community. Medical comorbidities, isolation, pain, or dementia are among the risks for a late-onset first episode, and risk increases after age 75 (Blazer, 2002; Federal Interagency Forum, 2006). Seventy-five percent of older adults with depressive symptoms are seen by a primary care provider (Boswell & Stoudemire, 1996). Routine screening for depressive disorders in older adults in primary care settings has been endorsed by several professional organizations, notably the Institute of Medicine, American Association of Colleges of Nursing, and the Hartford Institute Nurses Improving Care for Healthsystem Elders faculty (Kurlowicz, 1997). Nevertheless, depression in older adults remains unrecognized and under-treated (Hirschfeld et al., 1997).

The eight instruments discussed in this targeted review were chosen due to their prevalence in recent literature on older adult depression, their usage in clinical settings with older adult populations, and their operational characteristics in older adult populations. The Geriatric Depression Scale (GDS; Yesavage et al., 1983) and the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977) currently are cited most often in the literature regarding assessment and recognition of

depression in older adults in clinical and research settings. The Hamilton Depression Scale (HAM-D) (Hamilton, 1960, 1967), and the Beck Depression Inventory (BDI) (Beck, Ward, & Mendelson, 1961), have been included because of their long standing usage by mental health professionals in clinical research. The Brief Patient Health Questionnaire-9 (PHQ-9) (Kronke, Spitzer, & Williams, 2001) and the Cornell Scale for Depression in Dementia (CSDD) (Alexopoulous, Abrams, Young, & Shamoian, 1988) are included because they are increasingly being mentioned in the older adult literature.

Geriatric Depression Scale (GDS)

The Geriatric Depression Scale (GDS; Yesavage et al., 1983) was created to detect probable cases of depressive disorders in older adults (see The Appendix). It is the only instrument specifically designed for this population. It asks 30 questions concerning feelings and behaviors within the last week. With its yes-no, forced choice format, the GDS is a simple instrument without gradation or frequency choices. Its original 30-item questionnaire has had subsequent iterations, decreasing the number of items to 15, 10, and even 5 items. It was designed to be self-administered, but is sometimes administered with the help of another person who does not need to be a professional. All instruments discussed herein are listed in comparitive table format as an appendix following this article.

The items on the scale do not rephrase the DSM criteria for a depressive episode; they were derived from a list of statements that the investigators believed to be related to late-life depression from their clinical research experience. When the scale was pilot-tested, the 30 items chosen were those that had the highest correlation to the total score out of 100 items. Items that had a somatic component were dropped based on their low correlation to the total score. Those 30 items comprising the GDS were found to have a sensitivity of 84% and a specificity of 95% with a cut-score of 11 or higher for detecting depression from a sample that included both normal controls and persons already diagnosed as being depressed (Yesavage et al., 1983).

Most depression scales include multiple items that elicit responses regarding somatic symptoms that often accompany depression. The GDS does not. In screening older persons, somatic complaints can be ambiguous. Their inclusion may lead to false positives in this population as these individuals have a high probability of chronic illness comorbidity. This does not imply that questions concerning problems with sleep and appetite are not important to diagnosis and treatment; but rather

that complaints of pains or abdominal problems may likely be more prevalent in older adults due to concurrent illnesses or medications. These confounds might diminish specificity. In confirming or monitoring adiagnosis, somatic aspects, including sleep, pain, appetite, and weight changes, are essential.

The original 30-item scale has remained unchanged since 1983 and has become the most frequently chosen by clinicians and researchers in geriatric psychiatry, in inpatient and outpatient studies, and is specifically recommended by the Institute of Medicine. This original 30-item GDS was superior to the CES-D or shorter versions of the GDS in detecting minor depressive disorders in primary care settings (Lyness et al., 1997). The original form is robust in detecting depression in nursing home residents, including those with a moderate degree of dementia when 11 was used as the cut-point (Jongenelis et al., 2005). (In this study, moderate dementia was classified by a score of 15 or greater on the Mini Mental State Exam (MMSE), a commonly used cognitive screening measure (McGivney, Mulvihill, & Taylor, 1994).) The importance of case-finding for depression in nursing home residents is critical. Rates of major depressive disorders in residents have been estimated at greater than 10%. The prevalence of significant depressive symptomatology, including minor depression, has been found at rates ranging from 30% (Alexopolous et al., 2001) to 44% in one study of urban nursing home residents (Teresi, Abrams, Holmes, Ramirez, & Eimicke, 2001).

Nevertheless, results from other clinical research with persons with Alzheimer's disease have been mixed. Parmelee, Katz, and Lawton (1989) found the GDS to correlate well with observational studies for depressive behaviors in institutionalized elders until the level of dementia interfered with comprehension. In another study that same year, the GDS was not deemed valid for those persons whose MMSE score was 25 or lower (Kafonek, Ettinger, & Roca, 1989).

Geriatric Depression Scale-15 (GDS-15; GDS-Short Form)

Although the GDS-15 (a subset of the 30 original items) has come into favor and has reasonable sensitivity and specificity when a cut-point score of 6 or 7 is used, it does not seem to be as robust as the original in minor depression detection, has not had the same usage, and has demonstrated a slightly lower sensitivity and specificity when compared to the 30-item scale (Cwikel & Ritchie, 1988). A recent, large scale study of community-based older adults supported the GDS-15 as having robust internal reliability, construct validity, and operational characteristics for the screening of community-dwelling, cognitively intact older adults

(Friedman, Heisel, & Delavan, 2005). Following a systematic and more exhaustive review of the literature on screening in primary care (Watson & Pignone, 2003), the 15-item version was also recently recommended for the screening of older adults in primary care settings. When an MMSE score is 15 or greater, the GDS-15 has been found to be superior in case-finding to the Minimum Data Set or to chart diagnosis in nursing home residents (Kerber, Dyck, Culp, & Buckwalter, 2005).

Center for Epidemiologic Studies Depression Scale (CES-D)

The Center for Epidemiologic Studies Depression Scale (CES-D) was designed to screen large numbers of community-dwelling adults for depressive symptoms sufficient to warrant further assessment (Radloff, 1977). It consists of 20 statements (e.g., "I was bothered by things that don't usually bother me"), each of which is to be answered according to frequency of occurrence within the past week. Responses are chosen from a Likert-type scale, ranging from rarely or none of the time (0) to most or all of the time (5).

The CES-D was designed for research and is frequently chosen in research studies. The tool has been validated for use in older adult populations. The instrument has shown mixed results in sensitivity and specificity depending upon the version used, designated cut-point, and population. The CES-D appears to have good utility in community-based settings but is less sensitive than the GDS for subsyndromal or minor depression (Lyness et al., 1997). For persons with cognitive impairment the number of choices and wording of the responses can be more difficult than the simple yes-no format of the GDS-15.

In a study of healthy community-dwelling old-old persons (mean age of 80 years) neither the CES-D nor the GDS was adequate in detecting major or minor depression in an affluent, educated group. With closer scrutiny it was determined that if cut-points had been set lower than the usual recommendations, either the GDS or the CES-D might have correctly identified approximately 80% of the suspected cases (Watson, Lewis, Kistler, Amick, & Boustani, 2004). This caveat may inform practice in similar settings of old-old persons.

Center for Epidemiologic Studies Depression Scale Short Form (CES-D Short Form)

Just as with the GDS, shorter variants of the CES-D have been derived from the original. The CES-D Short Form has been validated for use in older adult populations. The 10-item version demonstrated excellent

sensitivity and specificity for identifying major depression in clinics and hospital settings, while the GDS-15 was superior in nursing homes (Blank, Gruman, & Robison, 2004).

Hamilton Depression Scale (HAM-D)

The long-held psychiatric clinical standard, the Hamilton Depression Scale (HAM-D) was designed to be administered by a professional utilizing a semi-structured interview format (Hamilton, 1960; Hamilton, 1967). It became the standard for measurement in pharmaceutical clinical trials. It is not specific for older adult presentations of depressive syndromes and has multiple items related to somatic complaints, which may be confounded by the multiple chronic illnesses in older adults. Nevertheless, the HAM-D captures severity of symptomatology in multiple domains when used by a professional.

Beck Depression Inventory (BDI)

The Beck Depression Inventory (BDI) was intended to *measure* an existing depression, not to screen for or diagnose it, although it is often used as such (Beck, Ward, & Mendelson, 1961). A search on PsycInfo®, performed May, 2007, offered only six entries for the BDI and older adults, none more recent than 2001.

Brief Patient Health Questionnaire-9 (PHQ-9)

The Patient Health Questionnaire-9 is a self-administered version of the Primary Care Evaluation of Mental Disorders (PRIME-MD), a diagnostic scale for mental health disorders (Kroenke, Spitzer, & Williams, 2001). The Brief Patient Health Questionnaire-9 (PHQ-9) is a subset of this larger instrument. Its 9 items derived specifically from DSM-IV criteria for major depressive disorders. Early reliability and construct validity studies were carried out in primary care and in women's health outpatient settings. Items on this self-administered scale are rated by prevalence in the past 2 weeks, from 0 to 3 (nearly every day). It has been validated as a measure of severity and serves as a screening instrument for major depressive disorders. This dual utility of the PHQ-9 coupled with its psychometric rigor led to several recommendations that the PHQ-9 be the case-finding tool for use in primary care (Nease, Klinkman, & Aikens, 2006).

The PHQ-9 was chosen as a measurement instrument for older adults in the National Institute of Mental Health-sponsored IMPACT study of collaborative care management of depression in older adult primary

care (Unützer et al., 2002) and endorsed by the Geriatric Mental Health Foundation for use as a screening tool for older adults in primary care (American Association for Geriatric Psychiatry [AAGP], 2004). The PHQ-9 also was recently validated as a robust screening tool for post-stroke depression, "similar or superior to" other screening tools in this population (Williams et al., 2005, p. 638). However, in a comparison study with several other screening tools, the PHQ-9's superiority was only in identifying major depressive disorders. It did not fare as well in women or older adult subgroups and was less successful in detecting dysthymia (Henkel et al., 2004). Cut-points have been segmented in categories of severity; there are no specific guidelines for any variation in cut-points for older adult patients although this may be coming given the recent endorsements of the PHQ-9 by the aforementioned groups.

Cornell Scale for Depression in Dementia (CSDD)

Most of the tools previously mentioned lose operational characteristic strengths in the presence of more than minimal dementia. Nevertheless, the co-occurrence of depression in persons with dementia of the Alzheimer's type was recently found to range from 22–54% in a multi-site study (Zubenko et al., 2003). Rates of co-occurrence of depression with vascular dementia are estimated to be higher (Sultzer, Levin, Mahler, High, & Cummings, 1993). An instrument for assessing for depression in persons with dementia is the Cornell Scale for Depression in Dementia (CSDD; Alexopoulos, Abrams, Young, & Shamoian, 1988). This instrument depends on direct observation for all or a large part of the assessment as would be necessary when language and cognition are impaired. It was designed for broad screening. Depression at any stage of a dementing illness contributes to distress, poorer functioning, agitation, or even psychosis in the later stages of dementia and should be considered in differential diagnosis (Sachs-Ericsson & Blazer, 2006).

LIMITATIONS

The limits of screening tools as an entity have been discussed throughout this paper. Like any tool, they can achieve different outcomes depending on their goodness-of-fit to the task as well as the skill and expertise required of the users. This review has not been exhaustive, but rather, aimed at balancing research findings with clinical relevance, with a tilt toward the clinical. The applications of screening tools in research versus practice settings are quite different but resemble the efficacy versus effectiveness definitions used in treatment studies.

We have attempted to report on evidence-based practices while valuing practice-based evidence.

In practice settings, evidence includes more than a patient's subjective responses to questions concerning mood, affect, and distress. Physical signs also are important empirical data. Observable facial expressions, weight loss, and psychomotor agitation or retardation may suggest a depressive disorder even when complaints of depression are denied (Blazer, 2002). Both primary care and specialty providers must guard against overdependence on subjective data when screening for depression in this population.

IMPLICATIONS FOR PRACTICE

Although there have been many calls for routine screening of older adults for mood disorders in primary practice, it is far from standard practice. Screening tools are not diagnostic, but they can improve the assessment and monitoring of certain mood spectrum disorders commonly seen in older adults. Nurses are employed in many settings where screening for mood problems could facilitate improved care: community health agencies, long-term care facilities, hospitals, mental health settings, outpatient surgery centers, and primary care offices. With the improvement of care and quality of life as ultimate goals, the recognition of problems is necessary but not sufficient. Palmer and Coyne (2003) and others (Hirschfeld et al., 1997; Valenstein, Kales, & Mellow, 1998) have voiced caveats regarding the inadequacy of simply screening for depression. They refer to the under-treatment of cases already identified and that screening without proper infrastructure for diagnosis, treatment, and extensive follow-up is not acceptable (Rubenstein et al., 2007). Nevertheless, improved recognition in primary care settings can provide help, sometimes dramatically.

Older adult males have the highest rate of completed suicides (Conwell & Thompson, 2008). Many of these persons have visited their primary care provider within a month prior to the act. Depression is the strongest risk factor for late-life suicide and for suicide's precursor, suicidal ideation (Bruce et al., 2004). Recognizing depressive symptoms is a first step toward preventing suicide in older adults and warrants specific inquiry concerning thoughts of self-harm. Each of the instruments reviewed includes items that address hopelessness, but not all address self-harm directly. Hopelessness has been shown to correlate with increased suicidal risk in the older adult population, and responses suggesting hopelessness, futility, and burden warrant additional assessment even when the total score is less than the cut-point.

Some primary care providers are comfortable with treating moderate depressive symptoms in older adults; ideally referral or structured follow-up programs will be available for more difficult presentations. Models for structured care have been tested and have supported the efficacy of collaborative, integrated mental health services and health care in older adults with depression, anxiety, and alcohol-related problems. Three large scale, randomized, controlled intervention studies, using the acronyms: IMPACT, Improving Mood-Promoting Access to Collaborative Treatment (Unützer et al., 2002), PRISM-E, Primary Care Research in Substance Abuse and Mental Trial for the Elderly (Bartels et al., 2004; Oslin et al., 2006), and PROSPECT, Prevention of Suicide in Primary Care Elderly: Collaborative Trials, (Bruce et al., 2004) have provided models for older adult mental health services that are integrated with primary care. The IMPACT and PRISM-E studies utilized mental health professionals, including psychiatric mental health advanced practice nurses, either in-house or near the primary care provider's office to see patients promptly, treat them, and collaborate with the primary care provider. Indeed, a secondary analysis from the IMPACT study specifically reported high patient satisfaction with psychiatric mental health advanced practice nurses (Saur et al., 2007). Such models, when coupled with better recognition, have the potential for changing the horizon for mental health services delivery and improving outcomes for older adults in primary care settings.

However, until delivery systems change, screening tools have utility. There are still the usual routes of referral if a practitioner recognizes a mental health problem. As such, the adequacy of any screening tool depends upon the purpose, setting, and questions being asked. Tools that were devised for large epidemiological projects were not intended for individual, longitudinal measurement and vice versa. When the unit of study is the individual patient, the individual items take on more import, and the patient's baseline score becomes more informative than normative scores for a certain population. Nevertheless, adopting standard practices for the use of a tool keeps practitioners vigilant.

Future older adult cohorts, boomers and beyond, may be more psychologically minded and might freely initiate conversations concerning changes in mood or cognition; however, the insidious nature of changes in mood and cognition, or sometimes frank denial, will always warrant an active search by the clinician, regardless of the tools available in the future. Clinical significance, utility, and brevity will continue to be essentials for any instrument used in practice. Those tools described in this article may help mental health nurses increase their level of surveillance

and responsiveness to unspoken distress and unreported symptoms in today's older adult population.

REFERENCES

Alexopoulos, G. S., Abrams, R. C., Young, R. C., & Shamoian, C. A. (1988). Cornell Scale for Depression in Dementia. *Biologica Psychiatry, 23*(3), 271–284.

Alexopolous, G. S., Carpenter, D., Docherty, J. P., Katz, I. R., & Reynolds, C. F. (2001). The expert consensus guideline series. Pharmacotherapy of depressive disorders in older patients. *Postgraduate Medicine.* Special issue, Pharmacotherapy (Oct), 1–86.

American Association for Geriatric Psychiatry. (2004). Retrieved May 20, 2008, from http://www.treatmenthelps.org/

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*, (4th ed., text rev.). Washington, DC: Author.

Bartels, S. J., Coakley, E. H., Zubritsky, C., Ware, J. H., Miles, K. M., Arean, P. A. et al. (2004). Improving access to geriatric mental health services: A randomized trial comparing treatment engagement with integrated versus enhanced referral care for depression, anxiety, and at-risk alcohol use. *American Journal of Psychiatry, 161*, 1455–1462.

Beck, A. T., Ward, C. H., Mock, J., Erlbaugh, J., & Mendelson, M. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*(44), 53–63.

Blank, K., Gruman, C., & Robison, J. T. (2004). Case-finding for depression in elderly people: Balancing ease of administration with validity in varied treatment settings. *Journal of Gerontology: A Biolological Sciences/Medical Sciences, 59*, 378–384.

Blazer, D. G. (2002). *Depression in late life* (3rd ed.). New York: Springer.

Boswell, E. B., & Stoudemire, A. (1996). Major depression in the primary care setting. *American Journal of Medicine, 101*(6), 3S–9S.

Bruce, M. L., Ten Have, T. R., Reynolds, C. F., III, Katz, I. I., Schulberg, H. C., Mulsant, B. H., et al. (2004). Reducing suicidal ideation and depressive symptoms in depressed older primary care patients: A randomized controlled trial. *Journal of the American Medical Association, 291*, 1081–1091.

Conwell, Y., & Thompson, C. (2008). Suicidal behavior in elders. *Psychiatric Clinics of North America, 31*(2), 333–356.

Cwickel, J., & Ritchie, K. (1998). The Short GDS: Evaluation in a heterogeneous, multilingual population. *Clinical Gerontologist, 8*(2), 63–69.

Federal Interagency Forum. (2006). *Federal interagency forum on aging-related statistics. Older Americans update.* Washington, DC: U.S. Government Printing Office.

First, M. B., & Gibbon, M. (2004). *The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II)*. Hoboken, NJ: John Wiley and Sons.

Friedman, B., Heisel, M. J., & Delavan, R. L. (2005). Psychometric properties of the 15-item Geriatric Depression Scale in functionally impaired, cognitively intact, community-dwelling elderly primary care patients. *Journal of the American Geriatrics Society, 53*, 1570–1576.

Gallo, J., Fulmer, T., Paveza, G., & Reichel, R. (2000). *Handbook of geriatric assessment* (3rd ed.). Philadelphia: Jones and Bartlett.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23*(?), 56–62.

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6*, 278–296.

Henkel, V., Mergl, R., Kohnen, R., Allgaier, A. K., Moller, H. J., & Hegerl, U. (2004). Use of brief depression screening tools in primary care: Consideration of heterogeneity in performance in different patient groups. *General Hospital Psychiatry, 26*(3), 190–198.

Hirschfeld, R. M., Keller, M. B., Panico, S., Arons, B. S., Barlow, D., & Davidoff, F. (1997). The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. *Journal of the American Medical Association, 277*(?), 333–340.

Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult: Criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Archives of Internal Medicine, 159*, 1701–1704.

Jongenelis, K., Gerritsen, D. L., Kluiter, H., Derksen, M., Pot, A. M., Eisses, A. M. H., et al. (2005). Diagnostic accuracy of the original 30-item and shortened versions of the Geriatric Depression Scale in nursing home patients. *International Journal of Geriatric Psychiatry, 20*, 1067–1074.

Kafonek, S., Ettinger, W. H., Roca, R., Kittner, S., et al. (1989). Instruments for screening for depression and dementia in a long-term care facility. *Journal of the American Geriatrics Society, 37*(1), 29–34.

Kaplan, R. M., & Sacuzzo, D. P. (1989). *Psychological testing: Principles, applications and Issues.* New York: McGraw Hill.

Kerber, C. S., Dyck, M. J., Culp, K. R., & Buckwalter, K. (2005). Comparing the Geriatric Depression Scale, Minimum Data Set, and primary care provider diagnosis for depression in rural nursing home residents. *Journal of the American Psychiatric Nurses Association, 11*(5), 269–275.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PH-Q 9: Validity for a brief depression severity measure *Journal of General Internal Medicine, 16*, 606–613.

Kurlowicz, L. H. (1997). Nursing standard of practice protocol: Depression in elderly patients: In spite of its prevalence, associated negative outcomes, and good treatment response, depression in older adults is underrecognized, misdiagnosed, and undertreated. *Geriatric Nursing, 18*(5), 192–200.

Lichtenberg, P. A. (1999). *Handbook of assessment in clinical gerontology*. Hoboken, NJ: John Wiley and Sons.

Liebowitz, B., Pearson, J., Schneide, L., Reynolds, C., Alexopolos, G., Bruce, M., et al. (1997). Diagnosis and treatment of depression in late life: Consensus statement update. *Journal of the American Medical Association, 278*, 1186–1192.

Lyness, J. M., Noel, T. K., Cox, C., King, D. A., Conwell, Y., & Caine, E. D. (1997). Screening for depression in elderly primary care patients. *Archives of Internal Medicine, 157*, 449–454.

McGivney, S. A., Mulvihill, M., & Taylor, B. (1994). Validating the GDS depression screen in the nursing home. *Journal of the American Geriatrics Society, 42*(5), 490–492.

McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research, 4*, 293–307.

Nease, D. E., Jr., Klinkman, M. S., & Aikens, J. E. (2006). Depression case finding in primary care: A method for the mandates. *International Journal of Psychiatry in Medicine, 36*(2), 141–151.

Oslin, D. W., Grantham, S., Coakley, E. H., Maxwell, J., Miles, K. M., Ware, J. H. et al. (2006). PRISM-E: Comparison of integrated care and enhanced specialty referral in managing at-risk alcohol use. *Psychiatric Services, 57*(7), 954–958.

Palmer, S. C., & Coyne, J. C. (2003). Screening for depression in medical settings: Pitfalls, alternative and revised priorities. *Journal of Psychosomatic Research, 54*, 279–287.

Parmelee, P. A., Katz, I., & Lawton, M. P. (1989). Depression among institutionalized aged: Assessment and prevalence estimation. *Journal of Gerontology, 44*(2), M22–29.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Rubenstein, L. Z., Trinidad Hoyl, M., Harker, J. O., Alessi, C. A., Josephson, K. R., & Pietruszka, F. M. (2007). A randomized trial of a screening, case finding, and referral system for older veterans in primary care. *Journal of the American Geriatrics Society, 55*(2), 166–174.

Sachs-Ericsson, N., & Blazer, D. G. (2006). Depression and anxiety associated with dementia. In M. E. Agronin & G. J. Maletta (Eds.), *Principles and practice of geriatric psychiatry* (pp. 591–603). Philadelphia: Lippincott, Williams & Wilkins.

Saur, C. D., Steffens, D. C., Harpole, L. H., Fan, M. Y., Oddone, E. Z., & Unützer, J. (2007). Satisfaction with outcomes of depressed older adults with psychiatric clinical nurse specialists in primary care. *Journal of the American Psychiatric Nurses Association, 13*(1), 62–70.

Sultzer, D., Levin, H., Mahler, M., High, W., & Cummings, J. (1993). A comparison of psychiatric symptoms in vascular dementia and Alzheimer's disease. *American Journal of Psychiatry, 150*, 1806–1812.

Teresi, J., Abrams, R., Holmes, D., Ramirez, M., & Eimicke, L. (2001). Prevalence of depression and depression recognition in nursing homes. *Social Psychiatry & Psychiatric Epidemiology, 36*, 613–620.

Unützer, J., Katon, W., Callahan, C. M., Williams, J. W., Jr., Hunkeler, E., Harpole, L. et al. (2002). Collaborative care management of late-life depression in the primary care setting: A randomized controlled trial. *Journal of the American Medical Association, 288*, 2836–2845.

Valenstein, M., Kales, H., & Mellow, A. (1998). Psychiatric diagnosis and intervention in older and younger patients in a primary care clinic: Effects of a screening and diagnostic instrument. *Journal of the American Geriatrics Society, 46*, 1499–1505.

Watson, L. C., Lewis, C. L., Kistler, C. E., Amick, H. R., & Boustani, M. (2004). Can we trust depression screening tools in the oldest old? *International Journal of Geriatric Psychiatry, 19*, 278–285.

Watson, L. C., & Pignone, M. P. (2003). Screening accuracy for late-life depression in primary care: A systematic review. *Journal of Family Practice, 52*, 956–964.

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

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., et al. (1983). Development and validation of a Geriatric Depression Scale: A preliminary report. *Journal of Psychiatric Research, 17*(1), 37–49.

Zubenko, G. S., Zubenko, W. N., McPherson, S., Spoor, E., Marin, D. B., Farlow, M. R., et al. (2003). A collaborative study of the emergence and clinical features of the major depressive syndrome of Alzheimer's disease. *American Journal of Psychiatry, 160*, 857–866.

APPENDIX 1. Screening Instruments for Depression in Older Adults

Instrument	# Of Items	Rating & Scoring	Operational Characteristics	Strengths & Weaknesses	Reasons To Select The Instrument
Geriatric Depression Scale (GDS)	30	• Self-rated • Dichotomous (yes/no) • Some items are reverse scored	• 92% sensitivity • 89–95% specificity at a cut-point of 11 • 85% sensitivity and 69% specificity for minor depression in one nursing home study	• Specifically created to detect depressive disorders in older adults • Not based directly on DSM criteria but from statements the authors believed to be related specifically to depression in later life • Excludes somatic items • Has no suicide item • Robust in detecting depression in persons in institutional as well as community settings, including in those with a moderate degree of dementia	• Ease of administration • Simple yes/no answers • Superior to the CES-D or shorter versions of the GDS in the presence of mild to moderate dementia • Most frequently used instrument for older adults
Geriatric Depression Scale-15 (GDS-15; GDS-Short Form)	15	Same as original	• High correlation ($r = .84$ and $p = .001$) with the GDS • 92% sensitivity and 81% specificity with a cut-point of 5 in a primary care older adult screening	• Has also been shortened to 10 and 5 item variants • Shorter forms lose sensitivity in the presence of dementia, and generally lose specificity as compared to the long form. • Not as robust as the original in detecting minor or subsyndromal depression	• Quick to administer, taking 5–7 minutes • Has had increased usage in research recently

(Continued on next page)

APPENDIX 1. Screening Instruments for Depression in Older Adults (*Continued*)

Instrument	# Of Items	Rating & Scoring	Operational Characteristics	Strengths & Weaknesses	Reasons To Select The Instrument
Center for Epidemiological Studies Depression Scale (CES-D)	20	• Self-rated • Likert scale	Sensitivity and specificity in older adult populations (See text.)	• Based on DSM-II criteria. • Less sensitive than the GDS for subsyndromal or minor depression. • Could be more difficult for persons with cognitive impairment to complete.	• Originally designed to screen large numbers of community-dwelling adults for depressive symptoms • Good utility in community-based settings, but one study reported that the CES-D may not consistently measure symptoms at the item level in older adults and ethnically diverse primary care patients.
Center for Epidemiological Depression Scale Short Form (CES-D Short Form)	10	Same as above	• Sensitivity and specificity in clinics, home, and hospital settings • Reliability statistics of the 10-item CES-D were found to be comparable to those reported for the original CES-D. • In one sample of older adults, the sensitivity of the CES-D was 100%, specificity was 93%, and positive predictive value was 38% for major depression.	Same as above	• Brevity and increased usage • Not as sensitive for minor depressive disorders as for major depressive disorders

Instrument					
Hamilton Depression Scale (HAM-D)	21	• Interviewer-rated • Requires training in interviewing and scoring • Requires 15 to 20 minutes	Few studies testing psychometrics with older adults	• Items about somatic complaints may be subject to confounds due to multiple chronic illnesses in older adults. • Not suited for primary care settings due to training requirement	• Still prevalent in research settings and has been used to identify depression in dementia • Burdensome for screening purposes in nonpsychiatric settings
Beck Depression Inventory (BDI)	21	Self-rated	Few studies testing psychometrics with older adults	• Multiple somatic items • Not intended as a screening instrument	• Seldom chosen in clinical gerosychiatric settings (more often used in primary care settings and with younger adults)
Brief Patient Health Questionnaire-9 (PHQ-9)	9	• Self-rated • Cut-scores vary with diagnostic category and severity	• Sensitivity and specificity in an older adult sample (was a subset of a larger validation study) • Had good sensitivity and specificity with a post-stroke sample	• Developed specifically from the DSM-IV criteria • Can assess both diagnostic criteria and severity, but not as robust in older adults	• The tool of choice by Family Physician Journal • Used in IMPACT and PRISM-E studies • Available at: www.treatmenthelps.org
Cornell Scale for Depression in Dementia (CSDD)	19	• Observer-rated • Requires training to learn to use correctly	• Designed specifically for older adults with dementia • Incorporates information from the older adult, family members, caregivers, and health care records	• Requires skill to assess older adults with moderate and severe dementia	• Appropriate for use with older adults with dementia, especially moderate and severe dementia

Developed from American Association for Geriatric Psychiatry (2004), Alexopoulos et al. (1988), Henkel et al. (2004), Irwin, Artin, & Oxman (1999), Jongenelis et al. (2005), Lyness et al. (1997), and Williams et al. (2005).