Psychology and Aging
1997, Vol. 12, No. 2, 277–287


0882-7974/97/$3.00

# Center for Epidemiologic Studies Depression Scale (CES-D) As a Screening Instrument for Depression Among Community-Residing Older Adults

Peter M. Lewinsohn and John R. Seeley
Oregon Research Institute

Robert E. Roberts
University of Texas Health Science Center at Houston

Nicholas B. Allen
Oregon Research Institute

The efficacy of the Center for Epidemiologic Studies Depression Scale (CES-D) as a screener for clinical depression was examined in a sample of 1,005 community-residing adults (age range=50–96). Presence of a depressive disorder was determined by diagnostic interview. Analyses revealed that neither age, gender, cognitive impairment, functional impairment, physical disease, nor social desirability had a significant negative effect on the psychometric properties or screening efficacy of the CES-D. These results indicate that there was no significant degradation in the ability of the CES-D to screen for depression among community-residing elderly adults. This conclusion must be tempered by the fact that the sample did not include participants with the more disabling forms of cognitive or functional impairment and physical illness.



Depression is a common and often serious clinical problem among people of all ages and especially among those with acute and chronic medical conditions (Feinson, 1985; Kennedy, Kelman, & Thomas, 1990; Kinzie, Lewinsohn, Maricle, & Teri, 1986; Kitchell et al., 1982; Newmann, 1989; Rapp, Parisi, Walsh, & Wallace, 1988; Waxman, Carner, & Klein, 1984), many of whom are older adults. Ironically, studies have repeatedly shown that clinicians have difficulty detecting depression among elderly adults (Blazer & Williams, 1980; Bowers, 1990; Dessonville, Gallagher, Thompson, Finnell, & Lewinsohn, 1982; Hankin & Locke, 1982; Janowsky, 1982; Schuckit, Miller, & Hahlbohm, 1975; Schulberg et al., 1985; Zich, Attkisson, & Greenfield, 1990). For example, Rapp et al. found that only 9% of depressed patients admitted with physical illness were identified as depressed by hospital staff. Because of the underrecognition of depression among older adults, being able to screen effectively for possible depression would seem to be especially important (Katona, 1994). A major purpose of the present study was to evaluate the psychometric properties and the efficacy of a commonly used self-report depression scale—the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977)—with elderly adults when used as the first-level screener in a two-stage case identification procedure, the second stage being a diagnostic interview.

*Screening* is of central importance in many health-related fields, including psychopathology (Morrison, 1985). Its main purpose is to identify from a large group of people those who have an elevated probability of having the disorder under study. Screening tests like the CES-D are usually used as the first stage of the two-stage process (Diamond & Lilienfeld, 1962; Dohrenwend & Dohrenwend, 1982; Kendall, Hollon, Beck, Hammen, & Ingram, 1987; Lewinsohn & Teri, 1982; Morrison, 1985; Shrout & Fleiss, 1981; Shrout, Skodol, & Dohrenwend, 1986) in which those who score above a prespecified cutoff are identified as *putative cases*. The latter are given a diagnostic interview on the basis of which a definite diagnosis is established or ruled out. The screener should provide a fast, economical, and valid way of partitioning the study population into presumed well (negative on screening) and presumed ill (positive on screening) groups. Because of the associated complexities, a brief discussion of the ways to evaluate the efficacy of screeners is presented next.

In general terms, the efficacy of a screener is a function of its psychometric properties and of the prevalence, or base rate, with which the disorder occurs in the population to be screened. The relevant psychometric properties include *concurrent validity* (i.e., the degree of congruence between the screener and the diagnosis, which is typically evaluated with the correlation coefficient and with kappa; Cohen, 1960) and *reliability* (i.e., the test–retest and the internal consistency of the screener; Shrout & Fleiss, 1981). Test–retest reliability establishes the degree of stability of the measure over time, whereas internal consistency establishes the extent to which responses to several questions intended to refer to the same concept are consistent with one another. As pointed out by Shrout and Yager (1989),

Peter M. Lewinsohn, John R. Seeley, and Nicholas B. Allen, Oregon Research Institute, Eugene, Oregon; Robert E. Roberts, School of Public Health, University of Texas Health Science Center at Houston.

This research was partially supported by Grants MH33572, AG1449, MH35672, and MH41278 from the National Institute of Mental Health.

Correspondence concerning this article should be addressed to Peter M. Lewinsohn, Oregon Research Institute, 1715 Franklin Boulevard, Eugene, Oregon 97403-1983. Electronic mail may be sent via Internet to pete@ori.org.

screeners with high reliability (e.g., ones in which reliability is in the neighborhood of .90) can be shortened, sometimes by dropping as many as 75% of its items, without losing much of their sensitivity and specificity. On the basis of excellent psychometric properties that have been reported for the CES-D with general community populations, Shrout and Yager suggested that 5-item versions of the CES-D should be as sensitive and as specific as the full 20-item CES-D. This can be demonstrated mathematically, and empirical results have been consistent with this prediction (e.g., Fendrich, Weissman, & Warner, 1990; Hakstian & McLean, 1989; Roberts, Lewinsohn, & Seeley, 1991).

In practice, the efficacy of screening tests has been evaluated with a number of conceptually, somewhat different but empirically interrelated indexes. These include *sensitivity*—the ability of a test to give a positive finding when the person tested has or will develop the disorder under study—and *specificity*—the ability of the test to give a negative finding when the person tested is free from the disorder under study or is very unlikely to develop the disorder. Ideally, sensitivity and specificity should both be high, but they are inversely related. By shifting the cutoff up, one increases specificity but decreases sensitivity; by shifting the cutoff down, one accomplishes the opposite result. Sensitivity and specificity are usually expressed as percentages. In the case of a completely useless screener, they are both 50%; optimally they are both 100%. A related number is *hit rate* (HR), which is the percentage of participants correctly classified with given cutoff scores. A similar measure of diagnostic performance is the J Index (Youden, 1950), which ranges from zero to one; the index becomes unity only when both the false-positive and false-negative rates are zero.

Another way of evaluating the efficacy of a screener is through *receiver operating characteristic* (ROC) analysis. ROC analysis is a useful technique for showing what sensitivities and specificities can be achieved at different cutoffs on the screener (e.g., Hsiao, Bartko, & Potter, 1989; Mossman & Somoza, 1989; Murphy et al., 1987). By calculating *area under the ROC curve* (AUC), it is possible to compare different screeners and the efficacy of the same screener with different populations (Hanley & McNeil, 1982). On the basis of the ROC curve, it is also possible to select specific cutoff scores to maximize sensitivity and specificity. The AUC, the J Index, and the HR are closely related.

Screeners are also evaluated on the basis of their *positive predictive value* (PPV; the proportion of true positives among those who screen positive) or their *negative predictive value* (NPV; the proportion of true negatives among those screened negative). PPV and NPV are easily calculated from sensitivity and specificity if the prevalence of the disorder in question is known (Fleiss, 1981). Although sensitivity and specificity are not affected by the prevalence or base rate with which the disorder occurs in the population to be screened, PPV and NPV are strongly affected by the base rate (Dorn et al., 1996).

While there has been considerable interest in the assessment of depression in elderly adults (e.g., Katona, 1994), there have been relatively few studies conducted with the specific purpose of assessing depression among older adults (Allen et al., 1994; Blazer & Williams, 1980; Feinson, 1985; Gurland et al., 1983; Murrell, Himmelfarb, & Wright, 1983; Raymond, Michals, & Steer, 1980), and many of these studies have involved small and often severely ill, diagnostically difficult, or both, samples (e.g., Deforge & Sobal, 1988; Koenig, Meador, Cohen, & Blazer, 1988; Mahard, 1988; Norris, Gallagher, Wilson, & Winograd, 1987; O'Hara, Kohout, & Wallace, 1985). As Ranshoff and Freinstein (1978) pointed out, the operating characteristics of a screener are influenced by the degree of severity, chronicity, and comorbidity of the cases and of the controls. For example, a screener might do better when the to-be-screened population includes severely, chronically, and comorbid individuals than when the diseased group is composed primarily of individuals with mild forms of the disorder. Similarly, the degree to which depression and other pathologic conditions are very low or completely absent in the control group will affect sensitivity and specificity and, consequently, PPV and NPV. Given that the existing studies differ greatly on the above-mentioned parameters, it is difficult to generalize from them. On the basis of a review of these studies, Katona (1994) concluded, "There is as yet no ideal solution to the problem of detecting depression in elderly subjects with coexistent physical illness and/or dementia" (p. 19).

The CES-D, the depression measure evaluated in the present study, is the most widely used self-report depression measure in community studies (Gotlib & Cane, 1989). Its psychometric properties in studies with younger adults and with adolescents in clinical (Craig & Van Natta, 1976; Weissman, Sholomskas, Pottenger, Prusoff, & Locke, 1977) and community samples (Comstock & Helsing, 1976; Roberts et al., 1991; Singer & Willett, 1991) have consistently been shown to be strong. There is considerably less information concerning the psychometric properties of the CES-D with older adults. As the reviews by Radloff and Teri (1986) and by Weiss, Nagel, and Aronson (1986) indicated, most of these studies did not sample elderly adults in sufficiently large numbers to permit reliable comparisons at the upper end of the age continuum. Indeed, we are aware of only one study with an elderly community sample in which both the CES-D and clinical diagnosis were used. This study reported a HR of 82% (Murrell et al., 1983). Because we expected the psychometric properties of the CES-D in elderly adults to be strong, we also expected that it would be possible to generate a short form consisting, for example, of only 5 items, with an efficacy comparable to that of the full-length (20-item) form.

In this article we also evaluate the potential impact of four factors that might reduce the efficacy of the CES-D with older adults. We examined the effects of concurrent physical disease, of functional impairment, of the tendency to respond to test items in a socially desirable way, and of cognitive impairment on CES-D screening efficacy. We hypothesized that each of these factors increases with advancing age and, in turn, that each negatively affects the efficacy of the CES-D.

A major potential threat to the validity, and consequently the efficacy of self-report depression scales like the CES-D with elderly adults, stems from the overlap between some of the psychological concomitants of physical disease, and of functional impairment, with the symptoms of depression. Some of the symptoms of depression such as loss of appetite, problems with sleeping, and weight loss are also experienced by people suffering from many medical diseases (Cohen-Cole & Kaufman,

1993; Evans & Katona, 1993; Hughes, DeMallie, & Blazer, 1993; Kennedy, Kelman, & Thomas, 1989). Consequently, there is the danger that symptoms that are due to physical diseases and their associated functional impairment may be mistakenly attributed to depression leading to overdiagnosis of depression. Conversely, symptoms actually caused by depression may be misinterpreted as indications of physical disease leading to underdiagnosis. This problem would be expected to reduce the efficacy of the CES-D for people with medical diseases of all ages, but because an elevation of physical diseases and of physical symptoms has consistently been found in older adults (e.g., Murrell et al., 1983; Schulberg et al., 1985; Zich et al., 1990), the confounding effects of physical disease should be especially influential in these adults.

Although physical health problems are more frequent in elderly adults, and relationships between depression and a wide variety of diseases have been reported (e.g., Creed & Ash, 1992; Finch, Ramsay, & Katona, 1992; Wells, Rogers, Burnam, & Camp, 1993), the potential importance of the relationship of disability and of functional impairment with depression has also been recognized (Blazer, Burchett, Service, & George, 1991; Gurland et al., 1983; Gurland, Dean, & Cross, 1980; Linn, Hunter, & Harris, 1980; Smallegan, 1989; Zeiss, Lewinsohn, Rohde, & Seeley, 1996). Consequently, we also examined the impact of functional impairment on the efficacy of the CES-D. In this study, functional impairment was used to refer to the loss of ability to take care of one's everyday needs independently in the areas of work, self-care, or important leisure activities. Although physical disease increases the likelihood of developing functional impairment, the latter can occur in the absence of any specific disease as when there are age-related decrements in sensory function. On the basis of the above-mentioned considerations, we expected the analyses to show that increased physical disease and functional impairment are associated with aging, and elevated CES-D scores are observed in those with physical disease or functional impairment.

Another potential threat to the validity, and consequently the efficacy, of the CES-D with elderly adults may stem from the aging process itself. To the extent that increasing age results in increased reporting of depression-related complaints, items on scales such as the CES-D would be expected to be less valid for the measurement of depression in older adults. Indeed, because of the presumed similarity between the symptoms of depression and the "normal" manifestations of aging, it has been suggested that depression is more difficult to diagnose in older adults (Blazer & Williams, 1980; Dessonville et al., 1982; Janowsky, 1982). Thus, older people might score higher on a scale of depressive symptomatology, not because they are actually more depressed, but because the aging process itself may increase the likelihood of reporting problems such as sleeplessness, loss of appetite, and concentration difficulties—all of which are also symptoms of depression. Similarly, psychomotor retardation is one of the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed; *DSM–IV*; American Psychiatric Association [APA], 1994) symptoms of depression, and related constructs, such as slowed thinking, increased reaction time, and somatic complaints, have been shown to be associated with age (Earles & Salthouse, 1995; Hale, Myerson, & Wagstaff, 1987; Levkoff, Cleary, & Wetle, 1987; Mellinger, Balter, & Uhlenhuth, 1985; Salthouse, 1996). Indeed, Bolla-Wilson and Bleecker (1989) suggested that the increased prevalence of somatic complaints associated with aging may be misinterpreted as representing greater depression in the elderly population. This problem has also been recognized by Zemore and Eames (1979) who, using the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), found that the higher depression scores of older adults in their sample disappeared when somatic items from the BDI were not included in the analysis.

Another factor that might reduce the efficacy of self-report measures of depression such as the CES-D is cognitive impairment, which is also more common in older adults. Even though serious cognitive impairment affects only a small subgroup of elderly patients, especially those that are suffering from Alzheimer's disease (Reifler, Larson, & Handley, 1982), some level of cognitive impairment appears to be associated with normal aging (Lewinsohn, Rohde, Seeley, & Fischer, 1991). At severe levels of cognitive impairment, confusion and disorientation will work against the patient's ability to provide valid information, even on relatively simple questionnaire items such as those of the CES-D. However, even milder levels of cognitive impairment may attenuate validity to the extent that items require memory (e.g., "During the past week I had crying spells") or require self-awareness (e.g., person may be unaware that they have certain symptoms) and are susceptible to being misunderstood. The analyses to be presented were conducted with the expectation that the efficacy and psychometric quality of the CES-D would be reduced in community-residing older adults who exhibit some cognitive impairment.

Finally, in this article we examine the impact of yet another potential threat to the efficacy of the CES-D; namely, the tendency to respond to test items in a socially desirable way. Concern about the possibility that response biases may negatively affect the validity of test responses has a long history (Bernreuter, 1933; Edwards, 1957; Ellis, 1946; Jackson & Messick, 1962). Among the response biases, social desirability has been seen by some as a major threat to accurate assessment. Social desirability has typically been measured with items that reflect desirable behaviors with a very low frequency (e.g., "I am always willing to admit it when I make a mistake") and undesirable behaviors with a high frequency of occurrence (e.g., "I sometimes feel resentful if I don't get my way"). A number of tests have been developed to measure respondent tendencies to give socially desirable responses. The most popular among these is the one developed by Crowne and Marlowe (1960). Recent work in this area by Paulhus and his associates (e.g., Linden, Paulhus, & Dobson, 1986) suggested that in addition to a conscious attempt to look good to others (impression management), social desirability scales measure a second component; namely, a less conscious tendency to look good to oneself (self-deception or denial). Results by Paulhus and his group indicated that the Marlowe–Crowne Social Desirability Scale (Crowne & Marlowe, 1960) measures both of these components of social desirability. As might be expected, the Marlowe–Crowne Scale has been shown to be correlated with measures of depression (e.g., Lewinsohn et al., 1994; Linden et al., 1986; Spielberger, Gorsuch, & Lushene, 1970).

On the basis of previous literature (Lawton, Whelihan, &

Belsky, 1980), we expected that there would be a correlation between age and social desirability. Because high social desirability would be expected to result in underreporting of socially undesirable behavior and overreporting of socially desirable behavior (Nunnally, 1978), we hypothesized that the validity (and consequently the efficacy) of the CES-D would be lower for respondents who ranked high on social desirability.

The analyses for this article were conducted sequentially. First, we computed the psychometric properties of the CES-D when used in different age and gender groups of older people. Next, we examined the relationship between each of our independent variables (physical disease, functional impairment, cognitive impairment, and social desirability) with age and CES-D, with the expectation of positive associations. We then compared the efficacy of the CES-D for groups differentiated on the basis of age, physical illness, functional impairment, cognitive impairment, and social desirability. In all of these analyses, gender was included as a main and interactive effect. Finally, we expected to be able to generate an abbreviated CES-D with psychometric properties comparable to those of the full CES-D.

## Method

### Participants

From a list of licensed drivers 50 years of age or older, letters were sent to 4,133 randomly selected individuals between 1982 and 1983 informing them about the study and that they would be individually contacted. Follow-up calls were made to a random sample of 2,662 of these individuals to provide more information and to encourage participation. Of those called, 1,554 (58%) agreed to participate (10% were either ill or had moved out of the area; 32% declined to participate). Because of logistical constraints, only 1,008 of the 1,554 participants who initially agreed to participate were interviewed between 1982 and 1984. The remaining 546 individuals completed the questionnaire but did not attend the diagnostic interview (see Rohde, Lewinsohn, & Seeley, 1990, for additional information). Rohde et al. compared the participants in this project with U.S. census data for the area and found that participants were somewhat better educated and more likely to be women than community averages. Older individuals were slightly more likely to discontinue participation; no other assessed demographic differences were associated with attrition in this study.

The analyses were based on the sample of individuals who participated in the diagnostic interview and who completed the CES-D ($N = 1,005$). The demographic characteristics were as follows: 58% ($n = 586$) were women; the mean age was 63.9 ($SD = 7.9$); 6% had not completed high school, 63% completed high school, 18% had a bachelor's degree, and 13% had a professional degree; 29% were employed, 11% were unemployed, and 60% were retired; and 77% were married, 9% were divorced, 13% were widowed, and 1% had never married. The following age groups were used for this study: 50–59 ($n = 322$), 60–69 ($n = 451$), and 70 or older ($n = 232$).

### Diagnostic Classifications

Diagnostic information was gathered by using the Schedule for Affective Disorders and Schizophrenia—Lifetime version (SADS–L; Endicott & Spitzer, 1978); diagnoses were initially based on criteria provided by the Research Diagnostic Criteria (RDC; Spitzer, Endicott, & Robins, 1978) and later revised by using *DSM–III–R* (APA, 1994) criteria. Eighty participants (8%) met current criteria for major depressive disorder, dysthymia, or both (45 participants were diagnosed with major depressive disorder, 23 with dysthymia, and 12 with major depressive disorder and dysthymia).

Diagnostic interviewers were carefully selected and trained, completing a thorough didactic and experiential course in diagnostic interviewing. Every interview was either audiotaped or videotaped, and approximately 20% were independently rated by reliability coders. Interrater agreement was evaluated by means of the kappa statistic (Cohen, 1960). Kappas for individual disorders in all samples were consistently greater than .80, comparable to those reported by others (Orvaschel, Puigh-Antich, Chambers, Tabrizi, & Johnson, 1982; Spitzer et al., 1978).

### Measures

*CES-D.* The level of depressive symptomatology at the time of the questionnaire administration and the diagnostic interview was assessed with the CES-D (Radloff, 1977). The CES-D has been shown to possess adequate psychometric properties with adults (Husaini, Neff, Harrington, Hughes, & Stone, 1980; Radloff, 1977) and with adolescents (Roberts, Andrews, Lewinsohn, & Hops, 1990).

*Social desirability response style.* Ten items from the Marlowe–Crowne Social Desirability Scale (Crowne & Marlowe, 1960) were administered to measure the tendency to respond to test items in a socially desirable manner. The $T_1$ (Time 1) coefficient alpha of this measure was .68, and test–retest (the mean $T_1$–$T_2$ interval was 2.4 years, $SD = 0.7$) reliability was .74, $p < .001$. The total scale score was trichotomized to form low (0–3, $n = 342$), medium (4–5, $n = 296$), and high (6+, $n = 349$) groups; 18 participants had missing or incomplete scale scores.

*Cognitive impairment.* A composite measure of cognitive impairment was developed by using the following laboratory measures. The Trail Making Tests A and B assessed speed and sustained concentration on a simple and on a more complex visual task (originally from *Army Individual Test Battery*, U.S. War Department, 1944). Performance times on the two tests were standardized and summed. Digit Forward test of the WAIS–R (Wechsler, 1981) was administered as a measure of immediate attention span. Forty multiple-choice items from the Vocabulary section of the Shipley–Hartford Test (Shipley, 1940) were administered as a measure of verbal intelligence. To test speech audition, we computed a speech discrimination score from a phonetically balanced list of 50 words presented at a comfortably loud conversational decibel level with a tape recorder (in which the participant was asked to repeat each word). After completing the speech discrimination test, participants were asked to recall as many of the words that had been presented as possible, and a long-term memory score was computed. After this procedure, each word from a speech audiometry test was presented in print and paired with a word that had not appeared on the list; participants were asked to choose the word they had heard previously, and a recognition memory score was computed. The laboratory measures were standardized and summed to create a composite measure of cognitive impairment. Three groups of approximately equal size were formed: low impairment ($n = 264$), medium impairment ($n = 282$), and high impairment ($n = 286$); for logistical reasons, laboratory data were not available for 173 participants.

*Functional impairment.* A composite measure of functional status was developed by using the following self-report items: a five-point scale assessing vision (normal without glasses or contacts, normal corrected with glasses or contacts, somewhat impaired even with glasses or contacts, partially blind, or totally blind), a five-point scale assessing hearing (excellent, normal, poor, partially deaf, or totally deaf), a four-point scale assessing disabilities other than vision or hearing (no other disabilities, disability more than 1 year ago, disability 6 months to 1 year ago, or disability within the last 6 months), a four-point scale assessing the impact of health problems on doing activities (not at all,

a little, moderately, or a great deal), and a five-point scale assessing walking mobility (normal, somewhat impaired, seriously impaired, need walker, or need wheelchair). The total impairment score was based on the sum of the standard scores of the items. The internal consistency was moderate (Cronbach's $\alpha = .59$), suggesting that the items did not measure a single construct. However, this seems reasonable because impairment on one of the items should not necessarily indicate impairment status on the other items. Thus, the composite scores should be thought of as the summation of functional disabilities across several domains. Three groups of approximately equal size were formed: low impairment ($n = 361$), medium impairment ($n = 325$), and high impairment ($n = 295$); 24 participants had incomplete or missing impairment data.

*Physical disease.* Participants completed a 30-item checklist indicating which, if any, physical diseases they were currently experiencing. Among the 30 diseases listed were arthritis, emphysema, thyroid disease, Parkinson's disease, multiple sclerosis, coronary disease, meningitis, and glaucoma. Of the diseases reported, arthritis predominated, with coronary problems the second most frequent. Three disease groups were formed: no diseases ($n = 425$), one disease ($n = 341$), and two or more diseases ($n = 220$); 19 participants had missing or incomplete checklists.

## Results

### *Psychometric Properties of the CES-D*

Data on the operating characteristics of the CES-D are presented in Table 1. Internal consistency and test–retest reliability are acceptable and comparable across gender and age groups. The mean CES-D score did not differ significantly between male and female participants or between the three age groups. Likewise, no significant gender or age differences were found for the prevalence rates on the basis of the CES-D cutoff points of 16 and 20; no Age × Gender interactions attained statistical significance.

### *Association of CES-D and Age With Cognitive Impairment, Functional Impairment, Physical Disease, and Social Desirability*

Pearson product–moment correlation coefficients of the CES-D with (a) the cognitive impairment composite score, (b) the functional impairment composite score, (c) the number of physical diseases, and (d) the social desirability scale were computed. The correlations were .02 ($p > .05$), .23 ($p < .001$), .13 ($p < .001$), and −.14 ($p < .001$), respectively. The correlations of these four variables with age were .25, .15, .15, and .14 (all $ps < .001$), respectively. Thus, all of the correlations were statistically significant and in the expected direction, except for the nonsignificant association between CES-D and cognitive impairment. The correlation between CES-D and age in the sample was −.02 ($p > .05$).

### *Association Between CES-D Items and Age*

The correlations between the 20 CES-D items and age were also computed. Only two correlations were found to be significant at $p < .01$. Older participants endorsed Item 4 ("I felt that I was just as good as other people"; $r = -.14$) and Item 18 ("I felt sad"; $r = -.10$) less frequently than did younger participants.

### *Screening Characteristics of the CES-D*

To evaluate the sensitivity and specificity of the CES-D at various cutoff points, we used the ROC analysis. Shown in Figure 1 is the ROC curve for the CES-D across the entire sample, with the values of the various cutpoints indicated. On the basis of the ROC analyses, the cutoff point that maximizes both sensitivity and specificity for the total sample is 12. As can be seen, the curve is substantially above the random ROC (line of no information), which represents the sensitivity and specificity of random guesses as to whether an individual is a case or not. The fact that the AUC is significantly greater than this line means that the null hypothesis (i.e., that the CES-D is providing no useful information) can be rejected.

Data on the comparative ability of the CES-D to detect clinically diagnosable cases within various groups as well as for the total sample are presented in Table 2. Data are presented separately by age, gender, cognitive impairment, functional impairment, and social desirability, the latter three dimensions be-

Table 1
*Psychometric Properties of the Center for Epidemiologic Studies Depression Scale*

| Characteristic | Total sample | Sex | | Age (in years) | | |
|---|---|---|---|---|---|---|
| | | Female | Male | 50–59 | 60–69 | 70+ |
| Reliability | | | | | | |
| $\alpha$ | .82 | .83 | .81 | .83 | .83 | .78 |
| $T_1$–$T_2$[a] | .52 | .51 | .55 | .49 | .54 | .54 |
| Distribution of scores | | | | | | |
| *M* | 8.33 | 8.67 | 7.87 | 8.73 | 7.83 | 8.74 |
| *SD* | 6.84 | 7.16 | 6.33 | 7.03 | 6.78 | 6.63 |
| Mode | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 3.00 |
| Range | 0–46 | 0–46 | 0–41 | 0–38 | 0–46 | 0–34 |
| Prevalence (%) | | | | | | |
| ≥16 | 13.73 | 15.04 | 11.90 | 15.22 | 11.09 | 16.81 |
| ≥20 | 7.56 | 8.72 | 5.95 | 8.70 | 6.87 | 7.33 |

[a] For the $T_1$–$T_2$ interval, $n = 738$, $M = 2.4$ years, and $SD = 0.7$. $T_1$ = Time 1; $T_2$ = Time 2.

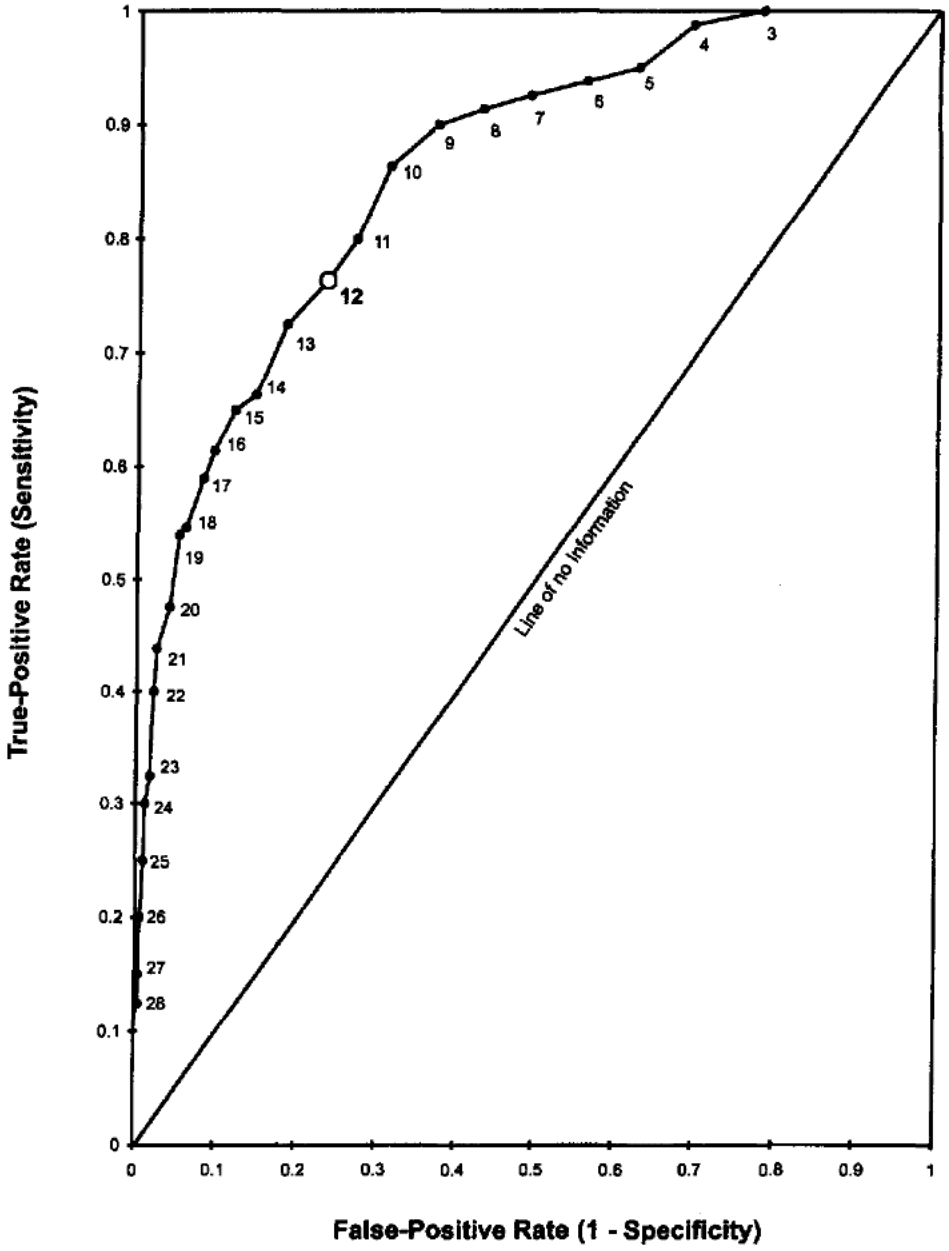


*Figure 1.* Receiver operating characteristic curve for the Center for Epidemiologic Studies Depression Scale across all participants.

ing trichotomized into approximately equal groups. The screening indexes are based on a cutoff point of 12, with the exception of the AUC and the correlation between CES-D and depression diagnosis. High concordance of the screening indexes between the various groups was found, and neither the J Index nor the AUC differed significantly between groups.

### *Abbreviated CES-D Screener*

Logistic regression analysis was used to select the best subset of CES-D items for the abbreviated CES-D Scale. A hierarchical model predicting depression diagnosis was computed. To control for effects of age and gender, we entered these two variables as the first block in the equation. The 20 CES-D items were entered as the second block in the equation in which the backward stepwise deselection method was used to remove items that were not significant at $p < .05$ (likelihood ratio statistic). The following 5 items were retained in the final solution: Item 3 ("I felt that I could not shake off the blues even with help from my family or friends"; Wald test = 11.70, $p < .001$), Item 6 ("I felt depressed"; Wald test = 19.24, $p < .001$), Item 8 ("I felt hopeful about the future"; Wald test = 8.18, $p < .01$), Item

Table 2
*Indexes of Screening Efficacy for the Center for Epidemiologic Studies Depression Scale by the Various Groups*

| Variable | Sensitivity | Specificity | NPV | PPV | HR | J (CI) | AUC (CI) | CES-D with diagnosis (*r*) |
|---|---|---|---|---|---|---|---|---|
| Age (in years) | | | | | | | | |
| 50–59 | 75 | 76 | 97 | 26 | 76 | .51 (.35–.67) | .88 (.80–.96) | .52 |
| 60–69 | 76 | 79 | 98 | 22 | 78 | .54 (.39–.69) | .86 (.78–.94) | .42 |
| 70+ | 80 | 73 | 98 | 17 | 74 | .53 (.32–.74) | .84 (.72–.97) | .38 |
| Sex | | | | | | | | |
| Female | 71 | 74 | 97 | 20 | 74 | .45 (.32–.58) | .83 (.76–.90) | .41 |
| Male | 84 | 80 | 98 | 25 | 81 | .64 (.51–.78) | .91 (.84–.98) | .51 |
| Cognitive impairment | | | | | | | | |
| Low | 84 | 77 | 98 | 27 | 77 | .61 (.45–.76) | .83 (.73–.93) | .52 |
| Medium | 79 | 80 | 99 | 17 | 80 | .59 (.37–.81) | .90 (.79–1.00) | .42 |
| High | 67 | 74 | 96 | 21 | 73 | .41 (.22–.59) | .88 (.79–.97) | .43 |
| Functional impairment | | | | | | | | |
| Low | 57 | 83 | 97 | 17 | 81 | .40 (.18–.61) | .84 (.73–.95) | .34 |
| Medium | 81 | 77 | 98 | 23 | 77 | .58 (.42–.74) | .88 (.79–.96) | .48 |
| High | 83 | 68 | 97 | 23 | 70 | .52 (.37–.66) | .86 (.77–.95) | .47 |
| Physical disease | | | | | | | | |
| None | 81 | 79 | 98 | 23 | 79 | .59 (.45–.74) | .90 (.83–.97) | .45 |
| One | 72 | 79 | 97 | 21 | 79 | .51 (.33–.69) | .85 (.75–.94) | .42 |
| Two or more | 74 | 68 | 96 | 21 | 68 | .41 (.22–.61) | .82 (.71–.92) | .45 |
| Social desirability | | | | | | | | |
| Low | 79 | 74 | 97 | 25 | 74 | .53 (.39–.68) | .84 (.76–.93) | .44 |
| Medium | 79 | 78 | 98 | 20 | 78 | .57 (.38–.76) | .88 (.78–.98) | .47 |
| High | 65 | 78 | 97 | 17 | 77 | .43 (.23–.63) | .85 (.75–.95) | .37 |
| Total | 76 | 77 | 97 | 22 | 77 | .53 (.43–.63) | .86 (.81–.92) | .45 |

*Note.* CES-D = Center for Epidemiologic Studies Depression Scale; NPV = negative predictive value; PPV = positive predictive value; J = Youden index J; AUC = area under the receiver operating characteristic (ROC) curve; CI = 95% confidence interval; HR = hit rate. Sensitivity, specificity, NPV, PPV, HR, and J are based on the CES-D cutoff score of 12 or greater, as this was the cutoff that provided the best balance between sensitivity and specificity according to the ROC curve analyses. A version of this table using the more conventional CES-D cutoff score of 16 is available by writing to Peter M. Lewinsohn.

10 ("I felt fearful"; Wald test = 11.63, $p < .001$), and Item 11 ("My sleep was restless"; Wald test = 4.59, $p < .05$).

The 5-item CES-D Scale had moderate internal consistency ($\alpha = .60$) and test–retest reliability ($r = .46$). The correlation between the 20-item scale and the 5-item scale was .80. On the basis of ROC analysis, the screening efficacy of the 5-item scale was found to be comparable to the total scale (AUC = .87;95% CI [confidence interval] = .82–.92). The correlation of the 5-item scale with the diagnosis of depression was .48. At the cutoff point of four, sensitivity = 80, specificity = 80, NPV = 98, PPV = 26, HR = 80, and J Index = .60 (95% CI = .51–.69); no significant differences by gender or by age groups were found. Thus, the 5-item CES-D performed as well as the 20-item CES-D in detecting diagnosable cases of depression.

## Discussion

Overall, the results support the use of the CES-D as a screener for depression among community-residing elderly people. The internal consistency, test–retest reliability, and validity were high for all sex and age groups, and the mean scores on the CES-D and the proportions above the cutoff score of 16 were comparable with those that have been observed in community studies of younger adults (Radloff & Locke, 1986). As expected, functional impairment, physical disease, and the tendency to answer test items in a socially desirable way were correlated with the CES-D and with age. However, neither age, gender, cognitive impairment, functional impairment, physical disease, nor social desirability had a significant deleterious effect on the psychometric properties and screening characteristics of the CES-D. Finally, we were able to construct a short (five-item) version with comparable screening properties to the full-length version.

A number of limitations of this study must be borne in mind when interpreting the results. First, because the participants in this study were community-residing members, this sample did not contain elderly people with very disabling forms of cognitive impairment, physical disease, functional impairment, and depression. Thus, although the conclusions drawn here can be applied to community-residing older adults, they may not be generalizable to elderly people whose illness requires that they reside in hospitals, nursing homes, or in other supported accommodation contexts. Indeed, Thompson, Heller, and Rody (1994) have presented evidence that depressed older adults may be especially likely to refuse participation in community surveys, suggesting that the range of depressive phenomena in our sample may be restricted.

Another shortcoming of the study is that it was a cross-sectional study, and, therefore, the age groups compared also represented different birth cohorts. As there is some evidence

to suggest that more recently born generations show a greater lifetime prevalence of depression than earlier born generations (e.g., Klerman et al., 1985; Lewinsohn, Rohde, Seeley, & Fischer, 1993), it is not ideal to have age and cohort effects confounded in this way. Nevertheless, the fact that there were no differences in the psychometric performance of the CES-D across these groups, despite the confluence of age and cohort effects, is further evidence of the robustness of this instrument.

The fact that the CES-D did not show a significant correlation with cognitive impairment in this sample is unexpected as depression scores have shown strong associations with test-based measures of cognitive impairment in previous studies with elderly participants (La Rue, Swan, & Carmelli, 1995; Rabbitt, Donlan, Watson, McInnes, & Bent, 1995), even when controlling for a range of associated demographic and health-related features. Indeed, research with elderly adults suffering from severe cognitive impairment, such as patients with Alzheimer's disease, has shown them to be unreliable information sources in the assessment of their own depression (Gilley et al., 1995). Thus, the restricted range of cognitive impairment in our sample may have reduced the impact of cognitive impairment on the properties of the CES-D.

Although the short form performed comparably to the full-length version in this sample, it needs to be recognized that the items for the short form were chosen for their efficacy in predicting diagnosis in this sample. It is important to cross-validate the short form proposed here on an independent sample of diagnosed older adults before it can be recommended as an alternative to the full-length CES-D for screening purposes. It is worth noting that there is little overlap between the items selected for short forms of the CES-D by other investigators (Kohout, Berkman, Evans, & Cornoni-Huntley, 1993; Krause & Liang, 1992) and those chosen by us. Only two of the items from the short form designed here are present on these other short forms: Items 6 ("I felt depressed") and 11 ("My sleep was restless"). This further emphasizes the need for cross-validation of the short form described in our study.

Our results with the CES-D are consistent with those that have been found in other studies. Stukenberg, Dura, and Kiecolt-Glaser (1990) also used ROC analysis to evaluate the relative efficacy of the BDI (Beck et al., 1961), the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983), and the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960) as screeners for depression in community-residing older adults. They found that the two questionnaire-based assessments (BDI and BSI) were comparable to the interview-based screener (HDRS). We are especially interested in the areas under the ROC curves for the three screeners in the Stukenberg et al. study (BDI = .82, BSI = .83, and HDRS = .85) that are comparable to those described in this study (range = .82–.91).

There are a number of interesting negative results in this study. One is that gender was not related to CES-D score. Amongst younger adults and adolescents, it is typical to find that female participants as a group score significantly higher on depression inventories than male participants (Nolen-Hoeksema, 1990). This preponderance of depressive symptoms amongst female participants, however, is usually not present before puberty and after the age of 65 (Nolen-Hoeksema, 1990). Most important, however, vis-à-vis the goals of our study, screening efficacy did not differ between female and male participants. Other researchers, however, have found gender differences in the sensitivity of self-report depression scales to detect clinical cases. Allen-Burge, Storandt, Kinscherf, and Rubin (1994) gave the Geriatric Depression Scale (GDS; Brink et al., 1982) and the BDI to a group of elderly psychiatric patients with unipolar depression. They found that many more men than women were "missed cases" (i.e., sensitivity was lower in elderly men). Because the sample in the study by Allen-Burge et al. consisted entirely of depressed cases, gender differences in specificity could not be examined.

The literature on the relationship between cognitive impairment and depression has produced equivocal results. Although we did not find a relationship, other recent studies have suggested that cognitive impairment is related to increased levels of depressive symptoms but not to diagnoses of clinical depression (Kay et al., 1985; Lindesay, Briggs, & Murphy, 1989). Most important, however, cognitive impairment did not have a deleterious effect on the screening efficacy of the CES-D in this sample—a finding that is consistent with that of other recent studies. Thus, Parmelee, Katz, and Lawton (1989) and Parmelee, Lawton, and Katz (1989) found that, except for the extremely demented, comparison of cognitively impaired versus intact respondents revealed that self-reports of depression were equally internally consistent and had equivalent correlations with observer ratings. Their findings suggest, as do ours, that except for the severely cognitively impaired, older adults are reliable informants regarding affective states.

Functional impairment and physical illness also did not adversely affect the psychometric and screening properties of the CES-D, even though functional impairments and physical disorders were associated with both depression and aging. These results corroborate those of Berkman and associates (Berkman et al., 1986) who found that physical disability with elderly adults does not degrade the validity of the CES-D.

The tendency to present information about oneself in a socially desirable manner also had no significant impact on the efficacy of the CES-D. In their literature review Gove and Geerken (1994) examined the extent to which social desirability moderates relationships between factors such as gender, race, education, income, age, marital status, and occupation with three indicators of mental health (psychiatric symptoms, self-esteem, and feelings of positive affect) and concluded that the social desirability response bias has very little impact on the relationships.

In their totality, the results presented in this article indicate that the efficacy of the CES-D as a screener for depression amongst community-residing elderly people is robust. We, therefore, recommend the use of the 20-item CES-D (in a two-stage process in which the CES-D is followed by a structured diagnostic interview for the high-scoring respondents) as a screener for depression amongst community-residing older adults. If the short form of the CES-D developed in this study is cross-validated, it too can be recommended as a useful alternative to the 20-item version for screening purposes. Despite these positive recommendations, the limitations of this study also suggest a number of important issues for future research. The potential impact of more severe levels of cognitive impairment, functional impairment, and physical disease needs to be examined,

as it is these individuals who pose the greatest challenge to clinicians attempting to detect depression. If these factors are shown to have an effect at higher levels of severity, where is the threshold? Another important direction for future research might be to examine the impact of age, physical health, functional impairment, and social desirability on relative screening efficacy of other self-report depression inventories, especially those that were designed for elderly populations (e.g., the GDS; Brink et al., 1982). Do they offer superior performance to the CES-D? Our study presents a benchmark, and ROC curve analyses can be a particularly effective tool for such comparisons.

## References


American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Allen, N., Ames, D., Ashby, D., Bennets, K., Tuckwell, V., & West, C. (1994). A brief sensitive screening instrument for depression in late life. *Age and Aging, 23,* 213–218.

Allen-Burge, R., Storandt, M., Kinscherf, D. A., & Rubin, E. H. (1994). Sex differences in the sensitivity of two self-report depression scales in older depressed inpatients. *Psychology and Aging, 9,* 443–445.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Berkman, L. F., Berkman, C. S., Kasl, S., Freeman, D. H., Jr., Leo, L., Ostfeld, A. M., Cornoni-Huntley, J., & Brody, J. A. (1986). Depressive symptoms in relation to physical health and functioning in the elderly. *American Journal of Epidemiology, 124,* 372–388.

Bernreuter, R. G. (1933). Validity of the personality inventory. *Personality Journal, 11,* 383–386.

Blazer, D., Burchett, B., Service, C., & George, L. C. (1991). The association of age and depression among the elderly: An epidemiologic exploration. *Journal of Gerontology, 46,* 210–215.

Blazer, D. G., & Williams, C. D. (1980). Epidemiology of dysphoria and depression in an elderly population. *American Journal of Psychiatry, 137,* 439–444.

Bolla-Wilson, K., & Bleecker, M. L. (1989). Absence of depression in elderly adults. *Journal of Gerontology: Psychological Sciences, 44,* P53–P55.

Bowers, W. A. (1990). Treatment of depressed inpatients: Cognitive therapy plus medication, relaxation plus medication, and medication alone. *British Journal of Psychiatry, 156,* 73–78.

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P., Adey, M., & Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical Gerontologist, 1,* 37–43.

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Cohen-Cole, S. A., & Kaufman, K. G. (1993). Major depression in physical illness: Diagnosis, prevalence, and antidepressant treatment (A ten year review: 1982–1992). *Depression, 1,* 181–204.

Comstock, G. W., & Helsing, K. J. (1976). Symptoms of depression in two communities. *Psychological Medicine, 6,* 551–563.

Craig, T. J., & Van Natta, P. A. (1976). Presence and persistence of depressive symptoms in patient and community populations. *American Journal of Psychiatry, 133,* 1426–1429.

Creed, F., & Ash, G. (1992). Depression in rheumatoid arthritis: Aetiology and treatment. *International Review of Psychiatry, 4,* 23–34.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349–354.

Deforge, B. R., & Sobal, J. (1988). Self-report depression scales in the elderly: The relationship between the CES-D and Zung. *International Journal of Psychiatry in Medicine, 18,* 325–338.

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13,* 595–605.

Dessonville, C., Gallagher, D., Thompson, L. W., Finnell, K., & Lewinsohn, P. M. (1982). Relation of age and health status of depressive symptoms in normal and depressed older adults. In S. Fleming (Ed.), *Essence* (Vol. 5, pp. 99–117). Atkinson, Ontario, Canada: College Press.

Diamond, E. L., & Lilienfeld, M. D. (1962). Effects of errors in classification and diagnosis in various types of epidemiological studies. *American Journal of Public Health, 52,* 1137–1144.

Dohrenwend, B. P., & Dohrenwend, B. S. (1982). Perspectives on the past and future of psychiatric epidemiology. *American Journal of Public Health, 72,* 1271–1279.

Dorn, L. D., Burgess, E. S., Susman, E. J., von Eye, A., DeBellis, M. D., Gold, P. W., & Chrousos, G. P. (1996). Response to CRH in depressed and nondepressed adolescents: Does gender make a difference? *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 764–773.

Earles, J. L., & Salthouse, T. A. (1995). Interrelations of age, health, and speed. *Journals of Gerontology: Series B. Psychological Sciences and Social Sciences, 50,* 33–41.

Edwards, A. L. (1957). *The social desirability variable in personality and research.* New York: Dryden Press.

Ellis, A. (1946). The validity of personality questionnaires. *Psychological Bulletin, 43,* 385–440.

Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry, 35,* 837–844.

Evans, S., & Katona, C. (1993). Epidemiology of depressive symptoms in elderly primary care attenders. *Dementia, 4,* 327–333.

Feinson, M. C. (1985). Aging and mental health: Distinguishing myth from reality. *Research on Aging, 7,* 155–174.

Fendrich, M., Weissman, M. M., & Warner, V. (1990). Screening for depressive disorder in children and adolescents: Validating the Center for Epidemiologic Studies Depression Scale for children. *American Journal of Epidemiology, 131,* 538–551.

Finch, E. J. L., Ramsay, R., & Katona, C. L. E. (1992). Depression and physical illness in the elderly. *Psychiatric Disorders in Late Life, 8,* 275–287.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Gilley, D. W., Wilson, R. S., Fleischman, D. A., Harrison, D. W., Goetz, C. G., & Tanner, C. M. (1995). Impact of Alzheimer's-type dementia and information source on the assessment of depression. *Psychological Assessment, 7,* 42–48.

Gotlib, I. H., & Cane, D. B. (1989). Self-report assessment of depression and anxiety. In P. C. Kendall & D. Watson (Eds.), *Anxiety and depression* (pp. 131–169). New York: Academic Press.

Gove, W. R., & Geerken, M. R. (1994). Response bias in surveys of mental health: An empirical investigation. *American Journal of Sociology, 82,* 1289–1317.

Gurland, B., Copeland, J., Kuriansky, J., Kellweher, M., Sharpe, L., & Dean, L. L. (1983). *The mind and mood of aging: Mental health problems of the community elderly in New York and London.* New York: Hawthorne Press.

Gurland, B., Dean, L., & Cross, B. (1980). The epidemiology of depression and dementia in the elderly: The use of multiple indicators in these conditions. In J. D. Cole & J. E. Barrett (Eds.), *Psychopathology in the aged* (pp. 37–62). New York: Raven Press.

Hakstian, A. R., & McLean, P. D. (1989). Brief reports: Brief screen for depression. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 139–141.

Hale, S., Myerson, J., & Wagstaff, D. (1987). General slowing of non-

verbal information processing: Evidence for a power law. *Journal of Gerontology, 42,* 131–136.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology and Neurosurgery, 23,* 56–61.

Hankin, J. R., & Locke, B. Z. (1982). The persistence of depressive symptomatology among prepaid group practice enrollees: An exploratory study. *American Journal of Public Health, 72,* 1000–1007.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under receiver operating characteristic (ROC) curve. *Radiology, 143,* 29–36.

Hsiao, J. K., Bartko, J. J., & Potter, W. Z. (1989). Diagnosing diagnoses: Receiver operating characteristic methods and psychiatry. *Archives of General Psychiatry, 46,* 664–667.

Hughes, D. C., DeMallie, D., & Blazer, D. G. (1993). Does age make a difference in the effects of physical health and social support on the outcome of a major depressive episode. *American Journal of Psychiatry, 150,* 728–733.

Husaini, B. A., Neff, J. A., Harrington, J. B., Hughes, M. D., & Stone, R. H. (1980). Depression in rural communities: Validating the CES-D Scale. *Journal of Community Psychology, 8,* 20–27.

Jackson, D. N., & Messick, S. (1962). Response styles and the assessment of psychopathology. In S. Messick & J. Ross (Eds.), *Measurement in personality and cognition.* New York: Wiley.

Janowsky, D. S. (1982). Pseudodementia in the elderly: Differential diagnosis and treatment. *Journal of Clinical Psychiatry, 43,* 19–25.

Katona, C. L. E. (1994). *Depression in old age.* New York: Wiley.

Kay, D. W. K., Henderson, A. S., Scott, R., Wilson, J., Pickerwood, D., & Grayson, D. A. (1985). Dementia and depression among the elderly living in the Hobart community: The effect of the diagnostic criteria on the prevalence rates. *Psychological Medicine, 15,* 771–788.

Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and recommendations regarding the use of the Beck Depression Inventory. *Cognitive Therapy and Research, 11,* 289–299.

Kennedy, G. J., Kelman, H. R., & Thomas, C. (1989). Hierarchy of characteristics associated with depressive symptoms in an urban elderly sample. *American Journal of Psychiatry, 146,* 220–225.

Kennedy, G. J., Kelman, H. R., & Thomas, C. (1990). The emergence of depressive symptoms in late life: The importance of declining health and increasing disability. *Journal of Community Health, 15,* 93–104.

Kinzie, J. D., Lewinsohn, P. M., Maricle, R., & Teri, L. (1986). The relationship of depression to medical illness in an older community population. *Comprehensive Psychiatry, 27,* 241–246.

Kitchell, M. A., Barnes, R. F., Veith, R. C., Okimoto, J. T., & Raskind, M. A. (1982). Screening for depression in hospitalized geriatric medical patients. *Journal of American Geriatrics Society, 30,* 174–177.

Klerman, G., Lavori, P., Rice, J., Reich, T., Endicott, J., Keller, M. B., Hirschfield, R., & Andreasen, N. C. (1985). Birth cohort trends in rates of major depressive disorder among relatives of patients with affective disorder. *Archives of General Psychiatry, 42,* 689–693.

Koenig, H. G., Meador, K. G., Cohen, H. J., & Blazer, D. G. (1988). Self-rated depression scales and screening for major depression in the older hospitalized patient with medical illness. *Journal of the American Geriatrics Society, 36,* 699–706.

Kohout, F. J., Berkman, L. F., Evans, D. A., & Cornoni-Huntley, J. (1993). Two shorter forms of the CES-D depression symptoms index. *Journal of Aging and Health, 5,* 179–193.

Krause, N., & Liang, J. (1992). Cross-cultural variations in depressive symptoms in later life. *International Psychogeriatrics, 4,* 185–202.

La Rue, A., Swan, G. E., & Carmelli, D. (1995). Cognition and depression in a cohort of aging men: Results from the Western Collaborative Group study. *Psychology and Aging, 10,* 30–33.

Lawton, M. P., Whelihan, W. M., & Belsky, J. K. (1980). Personality tests and their uses in older adults. In J. E. Birren & R. B. Sloane (Eds.), *Handbook of mental health and aging* (pp. 537–553). Englewood Cliffs, NJ: Prentice Hall.

Levkoff, S. E., Cleary, P. D., & Wetle, T. (1987). Differences in the appraisal of health between aged and middle-aged adults. *Journals of Gerontology, 42,* 114–120.

Lewinsohn, P. M., Roberts, R. E., Seeley, J. R., Rohde, P., Gotlib, I. H., & Hops, H. (1994). Adolescent psychopathology: II. Psychosocial risk factors for depression. *Journal of Abnormal Psychology, 103,* 302–315.

Lewinsohn, P. M., Rohde, P., Seeley, J. R., & Fischer, S. (1991). Age and depression: Unique and shared effects. *Psychology and Aging, 6,* 247–260.

Lewinsohn, P. M., Rohde, P., Seeley, J. R., & Fischer, S. (1993). Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *Journal of Abnormal Psychology, 102,* 110–120.

Lewinsohn, P. M., & Teri, L. (1982). Selection of depressed and nondepressed subjects on the basis of self-report data. *Journal of Consulting and Clinical Psychology, 50,* 590–591.

Linden, W., Paulhus, D. L., & Dobson, K. S. (1986). Effects of response styles on the report of psychological and somatic distress. *Journal of Consulting and Clinical Psychology, 54,* 309–313.

Lindesay, J., Briggs, K., & Murphy, E. (1989). The Grey's/Age Concern Survey: Prevalence rates of cognitive impairment, depression, and anxiety in an urban elderly community. *British Journal of Psychiatry, 155,* 317–329.

Linn, M. W., Hunter, K., & Harris, R. (1980). Symptoms of depression and recent life events in the community elderly. *Journal of Clinical Psychology, 36,* 675–682.

Mahard, R. E. (1988). The CES-D as a measure of depression mood in the elderly Puerto Rican population. *Journals of Gerontology, 43,* 24–25.

Mellinger, G. D., Balter, M. B., & Uhlenhuth, E. H. (1985). Insomnia and its treatment: Prevalence and correlates. *Archives of General Psychiatry, 42,* 225–232.

Morrison, A. S. (1985). *Screening in chronic disease.* New York: Oxford University Press.

Mossman, D., & Somoza, E. (1989). Assessing improvements in the Dexamethasone Suppression Test using receiver operating characteristic analysis. *Biological Psychiatry, 25,* 159–173.

Murphy, J. M., Berwick, D. M., Weinstein, M. C., Borus, J. F., Budman, S. H., & Klerman, G. L. (1987). Performance of screening and diagnostic tests: Application of receiver operating characteristic (ROC) analysis. *Archives of General Psychiatry, 44,* 550–555.

Murrell, S. A., Himmelfarb, S., & Wright, K. (1983). Prevalence of depression and its correlates in older adults. *American Journal of Epidemiology, 117,* 173–185.

Newmann, J. P. (1989). Aging and depression. *Psychology and Aging, 4,* 150–165.

Nolen-Hoeksema, S. (1990). *Sex differences in depression.* Palo Alto, CA: Stanford University Press.

Norris, J. T., Gallagher, D., Wilson, A., & Winograd, C. H. (1987). Assessment of depression in geriatric medical outpatients: The validity of two screening measures. *Journal of the American Geriatrics Society, 35,* 989–995.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: MacGraw-Hill.

O'Hara, M. W., Kohout, F. J., & Wallace, R. B. (1985). Depression among the rural elderly: A study of prevalence and correlates. *The Journal of Nervous and Mental Disease, 173,* 582–589.

Orvaschel, H., Puig-Antich, J., Chambers, W. J., Tabrizi, M. A., & Johnson, R. (1982). Retrospective assessment of prepubertal major depression with the Kiddie-SADS-E. *Journal of the American Academy of Child Psychiatry, 21,* 392–397.

Parmelee, P. A., Katz, I. R., & Lawton, M. P. (1989). Depression among institutionalized aged: Assessment and prevalence estimation. *Journals of Gerontology, 33,* 22–24.

Parmelee, P. A., Lawton, M. P., & Katz, I. R. (1989). Psychometric properties of the Geriatric Depression Scale among the institutionalized aged. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 331–338.

Rabbitt, P., Donlan, C., Watson, P., McInnes, L., & Bent, N. (1995). Unique and interactive effects of depression, age, socioeconomic advantage, and gender on cognitive performance of normal healthy older people. *Psychology and Aging, 10,* 307–313.

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.

Radloff, L. S., & Locke, B. Z. (1986). The community mental health assessment survey and the CES-D scale. In M. M. Weissman & J. K. Myers (Eds.), *Community surveys of psychiatric disorder* (pp. 177–189). New Brunswick, NJ: Rutgers University Press.

Radloff, L., & Teri, L. (1986). Use of the Center for Epidemiologic Studies Depression Scale with older adults. *Clinical Gerontologist, 5,* 119–136.

Ransohoff, D. F., & Feinstein, A. R. (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *New England Journal of Medicine, 299,* 926–930.

Rapp, S. R., Parisi, S. A., Walsh, D. A., & Wallace, C. E. (1988). Detecting depression in elderly medical inpatients. *Journal of Consulting and Clinical Psychology, 56,* 509–513.

Raymond, E., Michals, T., & Steer, R. (1980). Prevalence and correlates of depression in elderly persons. *Psychological Reports, 47,* 1055–1061.

Reifler, B. V., Larson, E., & Handley, R. (1982). Coexistence of cognitive impairment and depression in geriatric outpatients. *American Journal of Psychiatry, 139,* 623–626.

Roberts, R. E., Andrews, J. A., Lewinsohn, P. M., & Hops, H. (1990). Assessment of depression in adolescents using the Center for Epidemiologic Studies Depression Scale. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2,* 122–128.

Roberts, R. E., Lewinsohn, P. M., & Seeley, J. R. (1991). Screening for adolescent depression: A comparison of depression scales. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 58–66.

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1990). Are people changed by the experience of having an episode of depression? A further test of the scar hypothesis. *Journal of Abnormal Psychology, 99,* 264–271.

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review, 103,* 403–428.

Schuckit, M. A., Miller, P. L., & Hahlbohm, D. (1975). Unrecognized psychiatric illness in elderly medical–surgical patients. *Journal of Gerontology, 30,* 655–660.

Schulberg, H. C., Saul, M., McClelland, M., Ganguli, M., Christy, W., & Frank, R. (1985). Assessing depression in primary medical and psychiatric practices. *Archives of General Psychiatry, 42,* 1164–1170.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9,* 371–377.

Shrout, P. E., & Fleiss, J. L. (1981). Reliability and case detection. In J. Wing, P. Bebbington, & L. N. Robins (Eds.), *What is a case? The problem of definition in psychiatric community surveys* (pp. 117–128). London: Grant McIntyre.

Shrout, P. E., Skodol, A. E., & Dohrenwend, B. P. (1986). A two-stage approach for case identification and diagnosis: First-stage instruments. In J. E. Barrett & R. M. Rose (Eds.), *Mental disorders in the community: Progress and challenge* (pp. 286–303). New York: Guilford Press.

Shrout, P. E., & Yager, T. J. (1989). Reliability and validity of screening scales: Effect of reducing scale length. *Journal of Clinical Epidemiology, 42,* 69–78.

Singer, J. D., & Willett, J. B. (1991). Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin, 110,* 268–290.

Smallegan, M. (1989). Level of depressive symptoms and life stresses for culturally diverse older adults. *Gerontologist, 29,* 45–50.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State–Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research Diagnostic Criteria: Rationale and reliability. *Archives of General Psychiatry, 35,* 773–782.

Stukenberg, K. W., Dura, J. R., & Kiecolt-Glaser, J. K. (1990). Depression screening scale validation in an elderly, community-dwelling population. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2,* 134–138.

Thompson, M. G., Heller, K., & Rody, C. A. (1994). Recruitment challenges in studying late-life depression: Do community samples adequately represent depressed older adults? *Psychology and Aging, 9,* 121–125.

U.S. War Department. (1944). *Army individual test battery: Manual of directions and scoring.* Washington, DC: Author.

Waxman, H. M., Carner, E. A., & Klein, M. (1984). Underutilization of mental health professionals by community elderly. *The Gerontologist, 24,* 23–30.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised: Manual.* New York: Psychological Corporation.

Weiss, I. K., Nagel, C. L., & Aronson, N. K. (1986). Applicability of depression scales to the old person. *Journal of the American Geriatrics Society, 34,* 215–218.

Weissman, M. M., Sholomskas, D., Pottenger, M., Prusoff, B. A., & Locke, B. Z. (1977). Assessing depressive symptoms in five psychiatric populations: A validation study. *American Journal of Epidemiology, 106,* 203–214.

Wells, K. B., Rogers, W., Burnam, M. A., & Camp, P. (1993). Course of depression in patients with hypertension, myocardial infarction, or insulin-dependent diabetes. *American Journal of Psychiatry, 150,* 632–638.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3,* 32–35.

Zeiss, A. M., Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1996). Relationship of physical disease and functional impairment to depression in older people. *Psychology and Aging, 11,* 572–581.

Zemore, R., & Eames, N. (1979). Psychic and somatic symptoms of depression among young adults, institutionalized aged and noninstituionalized aged. *Journals of Gerontology, 34,* 716–722.

Zich, J. M., Attkisson, C. C., & Greenfield, T. K. (1990). Screening for depression in primary care clinics: The CES-D and the BDI. *International Journal of Psychiatry in Medicine, 20,* 259–277.

Received August 7, 1996
Revision received November 18, 1996
Accepted November 18, 1996 ■