Screening for Depression in the Older Adult

Criterion Validity of the 10-Item Center for Epidemiological Studies Depression Scale (CES-D)

Michael Irwin, MD; Kamal Haydari Artin, MD; Michael N. Oxman, MD

Background: The Center for Epidemiological Studies Depression Scale (CES-D) has been widely used in studies of late-life depression. While the CES-D is convenient to use in most settings, it can present problems for elderly respondents who may find the response format confusing, the questions emotionally stressful, and the time to complete burdensome. A briefer 10-item version has been proposed, but there are few data on its properties as a screening instrument.

Methods: The 10-item CES-D was administered in 2 studies. In study 1, a stratified sample of middle-aged depressed patients (n = 40) and comparison controls (n = 43) were administered the CES-D to determine an optimal cutoff score. In study 2, the accuracy of the CES-D optimal cutoff score was tested in a sample of adults older than 60 years (n = 68). Major depression diagnoses were derived from the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition, with consensus diagnoses using Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.

Results: Reliability statistics with the 10-item CES-D were found to be comparable to those reported for the original CES-D. Using an optimal cutoff score of 4 in study 1, the sensitivity of the 10-item CES-D was 97%; specificity, 84%; and positive predictive value, 85%. In the study 2 sample of older adults, the sensitivity of the CES-D was 100%; specificity, 93%; and positive predictive value, 38%.

Conclusion: The 10-item CES-D has excellent properties for use as a screening instrument for the identification of major depression in older adults.

Arch Intern Med. 1999;159:1701-1704

DEPRESSIVE symptoms and major depression are major public health problems in late life. Early diagnostic recognition of major depression in the primary care setting is needed if the morbidity associated with late-life depression is to be reduced. Consequently, there is considerable interest in the development of screening instruments that can be used to identify this often undiagnosed and undertreated disorder in the older adult.

Mulrow et al3 reported on studies of more than 15,000 primary care patients who had received screening with a case-finding instrument for depressive disorder. Several measures were found to have equivalent properties in terms of administration time and sensitivity and specificity for detecting major depression. However, findings from primary care patient groups may not apply to older persons. Differential underreporting of depressive symptoms by older adults might make some scales more accurate than others, depending on inventory emphasis of either somatic or affective symptoms.13

For editorial comment see page 1657

A few studies have now focused on older primary care patients and examined the case-finding characteristics of several self-report depression scales including the Center for Epidemiological Studies Depression Scale (CES-D),4–7 the Geriatric Depression Scale (GDS),4–7 the Beck Depression Inventory,8 and the Zung Depression Scale.9 Unfortunately, substantial methodological problems limit conclusions from many of these studies as recently reviewed by Lyness et al. Determination of caseness (the criterion standard) varied in validity or demonstrable reliability and/or the diagnos-
PATIENTS AND METHODS

In study 1, we evaluated the internal reliability, test-retest reliability, and optimal CES-D (10-item) cutoff score for the identification of depression in a stratified sample of depressed patients (n = 40) and controls (n = 43). The depressed patients and comparison controls were recruited by the University of California at San Diego Mental Health Clinical Research Center (MHCRC). Standardized recruitment strategies were used: placement of advertisements in newspapers and local campus newsletters, public announcements on the radio, and community educational outreach efforts. The patient populations and controls were identified from mostly middle-class San Diego neighborhoods and upper-class areas adjacent to La Jolla.

Both patients and controls were screened by MHCRC personnel and were then scheduled for a psychiatric diagnostic and medical evaluation. At that initial contact, informed consent was obtained as approved by the University of California at San Diego Human Studies Committee. To obtain a psychiatric diagnosis, all subjects were evaluated with the SCID with consensus diagnoses using DSM-IV. This interview was administered by either a trained psychiatric research fellow or psychiatric nurse who had undergone SCID reliability training provided by the MHCRC. Interview data were presented in a consensus conference of MHCRC diagnostic core psychologists, fellows, nurses, and other research personnel. Consensus diagnoses were assigned based on the SCID, physician investigator review of the medical chart, review of all available medical and psychiatric records, and SCID-based interview with family members and other resource informants when necessary. All depressed patients fulfilled DSM-IV criteria for major depressive disorder, current. Comparison controls had no lifetime psychiatric history and fulfilled criteria for never mentally ill. In addition to this psychiatric evaluation, all subjects underwent a medical history and physical examination by a psychiatric research fellow. Subjects who had an unstable medical condition or who had a physical disorder that might have led to a secondary depressive disorder were excluded. This exclusive approach was taken in study 1 to decrease the possibility that depressive symptoms were due to medical causes in our initial effort to determine a CES-D optimal cutoff score for the identification of primary major depression.

Following the initial psychiatric and medical evaluation, subjects completed the 10-item CES-D. Questionnaires were scored by personnel affiliated with this project and all MHCRC diagnostic staff were unaware of the CES-D scores. To determine reliability, a second 10-item CES-D was administered between 3 and 4 weeks later to all patients and controls. Because the CES-D is a state-dependent measure of depressive symptoms, administration of the CES-D in the depressed patients was completed before initiation of pharmacologic treatment by MHCRC psychiatrists. In study 2, we assessed whether the optimal 10-item CES-D cutoff score established in a stratified sample of depressed patients and controls was useful in the identification of major depression in a population of community-dwelling older adults. Sixty-eight subjects who were enrolled in local, university-affiliated primary care clinics provided informed consent for the present project. Subjects completed the 10-item CES-D and then underwent a psychiatric diagnostic assessment by research staff who were unaware of the CES-D scores. Consistent with study 1, the SCID was administered with consensus diagnosis using DSM-IV. However, in study 2, which focused on the diagnosis of major depression, only the mood module of the SCID was completed. Consensus diagnoses were assigned based on the SCID. An inclusive approach was used regarding symptoms, similar to that described by Lyness et al. In other words, symptoms were counted toward the criteria for depressive diagnoses without attempt to attribute them to psychiatric or medical causes.

In study 1, the reliability of the 10-item CES-D was determined by calculating Cronbach’s alpha in study 1. Test-retest reliability coefficients between time 1 and time 2 CES-D scores were obtained by Pearson correlation coefficients in the total sample in study 1. Parameters for criterion validity of the 10-item CES-D in study 1 and study 2 included sensitivity, specificity, and positive predictive value.

The measurement properties of the CES-D in older adults appear to be comparable to results generated in young and middle-aged adults. Reliability coefficients are high (0.85-0.91), and factor analyses replicate the 4-factor solution originally proposed by Radloff and Teri with only minor differences in patterns on the 4 CES-D subscales. There is no evidence of age-related increases in somatic scale scores with the CES-D, suggesting that physical disability among the elderly is not a major threat to the validity of the CES-D.

Despite the demonstrated utility of the CES-D in older adults, the original CES-D, a 20-item self-report index of depressive symptoms, can present problems for the older adult who may be unfamiliar with a multiple-item, forced-choice scale. Kohout found that elderly respondents were confused by the CES-D response for-
mat despite a prompt card listing the 4 options. Furthermore, the average administration time ranges from 7 minutes to as long as 10 to 12 minutes in many elderly respondents. To reduce administration time as well as to reduce response burden and to clarify response option, a briefer, simpler version of the CES-D has been proposed. 

A CES-D (10-item) version was developed at the Boston, Mass, site of the multisite project titled “Established Populations for Epidemiological Studies of the Elderly” (Table). Selected item subsets were based on factor analytic results reported by Radloff and Teri. Response options used a dichotomy that collapsed categories at both ends of the scale, resulting in a scale with an administration time of about 2 minutes. Data regarding the measurement properties of the 10-item version demonstrated that little precision is sacrificed with this abbreviated version. Factor analyses showed that the explained variance was somewhat greater for the 10-item version as compared with the original CES-D (66% vs 46.5%), and the pattern of loading was similar to the same 4-symptom dimensions in the 2 CES-D versions. Furthermore, reliability of the 10-item version (0.80, Cronbach α statistic) was only somewhat lower than that of the 20-item version (0.86). Finally, simulation analyses using Established Populations for Epidemiological Studies of the Elderly data generated at the Yale University (New Haven, Conn) site using the original 20-item version found a 0.88 correlation between 20-item and simulated 10-item scores, suggesting that it might be feasible to use regression equations to estimate 20-item scores.

Given that the briefer form retains an acceptable reliability and similar factor structure to the original CES-D along with an average time savings of about 5 minutes per interview and less burden of emotional stress, this study proposed to extend previous observations and to examine the screening characteristics and criterion validity of the CES-D 10-item version. Criterion standard diagnoses were based on the SCID interview, with consensus diagnoses using Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria. The first aim was to determine the test-retest reliability, operating characteristics, and optimal cutoff score of this scale in a sample of well-characterized patients with DSM-IV major depression and comparison controls who had no DSM-IV lifetime diagnosis (study 1). The second aim was to validate this optimal cutoff score in the identification of major depression in a sample of community-dwelling persons 60 years and older (study 2).

In study 1, 83 subjects were enrolled: 40 patients with a current diagnosis of DSM-IV major depression and 43 comparison controls with no lifetime history of any psychiatric disorder. Mean ± SD age was 44.9 ± 10.3 years in the depressed patients and 40.0 ± 12.8 years in the controls. Women comprised 50% (n = 20) of the depressed group and 42% (n = 18) of the control group.

The results from study 1 indicate that the 10-item CES-D has good internal consistency for the total sample (n = 83; Cronbach α = 0.92). Test-retest reliability of the CES-D was also high (Pearson r = 0.83; n = 83). The sensitivity, specificity, and positive predictive value of the 10-item CES-D were 97%, 84%, and 85%, respectively. A low cutoff score (≥4) was set to achieve a high rate of detection of depression in the depressed patients consistent with the proposed use of the CES-D as a screening instrument. Of the 40 depressed patients, 39 were correctly identified as having a major depression. One depressed subject had a score of 2 and was a false-negative. If the cutoff score was reduced to include that one depressive subject, the increase in sensitivity would be small compared with the loss in specificity.

Study 2 evaluated the ability of the CES-D to identify major depression in a sample of 68 community-dwelling older persons. The mean ± SD age of the sample was 72.0 ± 7.0 years and women comprised 52% (n = 35) of the study population. The CES-D optimal cutoff score (≥4) generated in study 1 was used. The sensitivity, specificity, and positive predictive value for the diagnosis of major depression in persons 60 years and older were 100%, 92%, and 38%, respectively. Of the 68 subjects, 3 had a current diagnosis of major depression and each of these subjects was correctly classified by the 10-item CES-D. The specificity of the scale was also high. The low positive predictive value indicates that the majority of those scoring above the threshold on the 10-item CES-D did not meet DSM-IV criteria for major depression.

To our knowledge, this is the first report of the criterion validity of the 10-item CES-D. The present findings generated in a stratified sample of depressed patients and nondepressed adults and in a population of older adults show that the briefer version of the CES-D has a reliability and validity comparable to that reported for the original CES-D. Second, the 10-item CES-D has excellent properties for use as a screening instrument in the identification of major depression in older adults. Using an optimal cutoff score (≥4) that was set in the stratified sample, the 10-item CES-D shows an excellent sensitivity and specificity in the screening of major depression in older adults. Diagnosis of major depression was made using a criterion standard approach, the SCID with consensus diagnosis using DSM-IV. Third, the operating characteristics of the 10-item CES-D in the older adult are comparable to the measurement prop-

### RESULTS

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I felt depressed</td>
<td>Yes</td>
</tr>
<tr>
<td>2. I felt that everything I did was an effort</td>
<td>Yes</td>
</tr>
<tr>
<td>3. My sleep was restless</td>
<td>Yes</td>
</tr>
<tr>
<td>4. I was happy</td>
<td>No</td>
</tr>
<tr>
<td>5. I felt lonely</td>
<td>Yes</td>
</tr>
<tr>
<td>6. People were unfriendly</td>
<td>Yes</td>
</tr>
<tr>
<td>7. I enjoyed life</td>
<td>No</td>
</tr>
<tr>
<td>8. I felt sad</td>
<td>Yes</td>
</tr>
<tr>
<td>9. I felt that people disliked me</td>
<td>Yes</td>
</tr>
<tr>
<td>10. I could not get going</td>
<td>Yes</td>
</tr>
</tbody>
</table>
properties of the well-validated original CES-D. As there is a considerable savings of time in the administration of the brief version, the 10-item CES-D may be very useful in the primary medical setting or in large-scale epidemiological studies that demand the briefest possible index of each health problem and risk factor. In contrast, where there is less concern about administration time or response burden, then change of depression over time may be more accurately assessed by a depression screening instrument such as the GDS, which is longer and includes more information about various depressive symptoms and their severity.

Several methodological issues might affect the interpretation of our data. The first concerns the possibility of sample bias. All subjects were patients in university-affiliated clinics and this sample may not be fully representative of community-dwelling adults or of primary medical clinic patients. Whether these results can be generalized to other settings must be defined empirically. Second, because of the low, 1-month prevalence of major depression and the sample size of older adults, the number of cases of major depression identified in the older adult samples was small. Third, the reliability and validity properties of this brief version of the CES-D were derived from the total score. There is some evidence that the total score on the CES-D may not be an accurate reflection of depression severity. Different populations do not respond in the same manner to individual items on the CES-D. For example, older adults have been found to be less likely to say that they felt blue and that people were unfriendly, suggesting that the use of these items on the 10-item CES-D may not adequately signal depression. This issue is particularly relevant to the present study since the appropriate cutoff score for the CES-D was determined in a population of middle-aged depressed subjects and controls and then applied to a second population of elderly adults more than 35 years their senior. Both the original and 10-item CES-D could be improved for use in older adults.

Despite these limitations, the present study used a well-validated diagnostic criterion standard and these data suggest that the 10-item CES-D is a reasonable screening tool for the identification of clinically significant major depressive disorders in older adults. Given the prevalence and morbidity associated with depression in late life along with ease of administration of the 10-item CES-D, clinicians should consider the routine use of this scale in older patients.

Accepted for publication January 7, 1999.

This work was supported by grant AA10215 from the National Institute on Alcohol Abuse and Alcoholism and grants MH18399 and M01RR0827 from the National Institutes of Health, Bethesda, Md; and Merck Research Laboratories UCSD Clinical Trial 960931.

We thank J. Christian Gillin, MD, John R. Kelsoe, MD, and Mark Rapaport, MD, for providing consensus diagnoses used in study 1.

Corresponding author: Michael Irwin, MD, Veterans Affairs Medical Center (Mail Code 116A), Department of Psychiatry, 3350 La Jolla Village Dr, San Diego, CA 92161 (e-mail: mirwin@ucsd.edu).

REFERENCES