Screening for Depression in Stroke Patients: The Reliability and Validity of the Center for Epidemiologic Studies Depression Scale

DAVID SHINAR, PH.D.,* CYNTHIA R. GROSS, PH.D.,† THOMAS R. PRICE, M.D.,‡ MARYANN BANKO, B.S.N.,‡ PAULA L. BOLDUC, B.S., and ROBERT G. ROBINSON, M.D.§

SUMMARY This study examined the inter-observer reliability and validity of the Center for Epidemiologic Studies Depression Scale (CES-D) as a measure of depressive symptomatology in stroke patients, and its utility as a screening tool for depression in this population. The CES-D Scale is a brief questionnaire originally designed for use in community surveys. Twenty-seven non-aphasic patients enrolled in the Stroke Data Bank at the University of Maryland were interviewed by a research nurse using the CES-D. On the same day, each patient was independently evaluated by a research assistant using a psychiatric battery for depression and measures of cognitive, physical, and social functioning. Forty-one percent (11/27) of the patients were depressed according to clinical criteria for major or minor depression. With a cutpoint corresponding to the upper (most severe) 20% in community surveys, the CES-D Scale picked up 73% (8/11) of the depressed patients. In this sample no nondepressed patient scored over 16 on the CES-D (no false positives). The CES-D Scale scores correlated significantly with the other depression measures (r = .57 to r = .82, p < .002) and did not correlate with the measures of cognitive, physical, or social functioning. Based on 24 patients who received a CES-D Scale score from both the nurse and the research assistant, inter-rater reliability was high (r = .76, p < .001). Thus, the CES-D was found to be reliable and valid as a screening tool for assessing depression in stroke patients.

PREVIOUS STUDIES have demonstrated that depression frequently follows stroke1-3 and that these post-stroke depressions, if untreated, may last many months.4,5 Therapy can ameliorate the symptoms of post-stroke depressions,6 and therefore it is important to find ways of quickly identifying stroke patients who are depressed. Since approximately 400,000 strokes occur in the United States each year,7 evaluation of all these patients by psychiatrists would be impractical. Consequently, a screening test for depression in stroke patients that is easy to administer, reliable, and valid is needed.

We propose that the Center for Epidemiologic Studies — Depression scale8 will satisfy these requirements, and be suitable to measure the depressive symptomatology of patients in the NINCDS Stroke Data Bank (SDB).9 This scale, widely known as the CES-D, is a brief self-report which can be easily administered to non-aphasic stroke patients by a research nurse. Although it has been extensively evaluated for its reliability and validity,10,11 the CES-D was designed for use in community surveys, and its appropriateness for stroke patients still needs to be demonstrated. This is because physical impairments due to stroke could invalidate responses to CES-D items such as “I felt that everything I did was an effort” which are intended to elicit symptoms of depression in otherwise healthy persons.

The purpose of the present study was to examine the inter-observer reliability, and the validity of the CES-D as a measure of depressive symptomatology for stroke patients, and its utility as a screening tool for depression in this population. Reliability was measured by comparing the scores on pairs of independently administered CES-D interviews. Concurrent validity was assessed by comparing CES-D scores with scores on a number of other depression measures, and discriminant validity was assessed by comparisons with measures of other factors including social functioning, cognition, and disability. Because CES-D scores were compared to several measures known to be differentially related to depression, the present evaluation can also be considered an assessment of construct validity. The utility of the CES-D as a screening tool for depression was evaluated by comparing CES-D performance (using an accepted cutoff score corresponding to the upper 20% in community surveys) with the clinical diagnosis of depression to determine false positive and false negative rates.

Method

Subjects

A consecutive series of 27 patients (11 men and 16 women) participating in the Stroke Data Bank (SDB) at the University of Maryland served as subjects in this study. They were all non-aphasic stroke patients and
either inpatients with a recent stroke or outpatients attending a Stroke Clinic between January and July, 1984. They ranged in age from 28 to 73 with 56 the median age.

Procedures

Patients were interviewed on one or more of the following three occasions: seven to ten days after stroke onset, three months after onset, or six months after onset. A total of 44 evaluations was obtained for the 27 patients. Patients were interviewed on the same day, first by the SDB research nurse and then by an experienced research assistant. The SDB nurse performed the functional evaluation of the SDB protocol which includes the CES-D. The research assistant administered a psychiatric battery of depression measures as well as cognitive, physical and social functioning questionnaires. Twenty-four of the 27 patients were also administered a CES-D by the research assistant who was blind to the nurse's findings. The second administration of the CES-D was done at the end of the second session so that it would not contaminate the other evaluations, and so that the time interval between its administration and the administration of the first CES-D would be maximized (to minimize recall of the first CES-D when responding to the second).

Measures

The Center for Epidemiologic Studies-Depression Scale (CES-D)15 was designed for use in household surveys as a self-rating scale to measure the severity of depressive symptomatology. It is a 20-item questionnaire (Appendix 1) investigating perceived mood and level of functioning within the past seven days. Scores range from 0–60, with higher scores indicating increasing severity of depression. Scores of 16 or higher are considered indicative of depression30 and correspond to the upper 20% of scores in a general population. A score of 20 or more has been recommended for people 55 years of age or older.3 In the present study, the interviewers read the CES-D items to the patient.

The psychiatric battery of depression measures included the Present State Examination (PSE)13 which was modified to assess effective mood and anxiety symptoms,3 the Hamilton Depression Scale31 and the Zung Depression Scale.15 The Johns Hopkins Functioning Inventory (JHFI)16,17 assessed degree of functional physical impairment, and the Mini-Mental State Examination18 assessed degree of cognitive impairment. The patient's level of social functioning was quantitatively assessed by the Social Scales Checklist and the Social Functioning Examination.19,20

A psychiatric diagnosis of depression was obtained using the symptoms elicited from the PSE and criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III).21 This conversion enabled classification of patients into one of three diagnostic categories: no depression, dysthymic (minor depression), or major depression. The method of conversion from PSE symptoms to DSM-III diagnosis has been previously described.3

Statistical Method

Two standard correlational techniques were employed. The Pearson product-moment correlation (r) was calculated whenever both measures were assumed to be on an interval scale. This is true of all the total scores. The Spearman rank correlation (rho) was calculated whenever one of the measures in the correlation had five or fewer discrete response categories. This was the case in all of the correlations involving individual items.

Inter-rater reliability was assessed by Kappa statistics,22 which measure the level of agreement, adjusted for agreement due to chance for items measured on a nominal scale. Kappa values range from −1.0 to 1.0, with values over .4 considered moderate agreement, values over .6 considered substantial agreement, and values over .8 considered excellent agreement.23

Data analyses were performed on the first evaluations of each patient (n = 24 for the inter-observer reliability of the CES-D, n = 27 for the validity assessments), as well as for the full data sets including follow-up evaluations on some patients (n = 34 for reliability, n = 44 for validity).

The analyses of the full data sets were performed for descriptive purposes only; it cannot be assumed that responses made by the same person on occasions three months apart were independent. In all instances the results for the full data set were similar to the findings from the first evaluations per person. Since independence of observations cannot be assumed for the full data set, this report contains only the results for the first visit data set.

Results

First patient evaluations consisted of six acute evaluations (7 to 10 days post-stroke), fifteen follow-up evaluations three months post-stroke, and six evaluations at six months post-stroke. Using symptoms derived from the first PSE, 26 percent of the patients had minor depression and an additional 15 percent had severe depression. Thus, 41 percent of the patients (11/27) were clinically depressed.

The Reliability of the CES-D

Two aspects of reliability were evaluated in this study: inter-observer reliability with the examiner-administration procedure, and internal consistency. Inter-observer reliability was determined by Pearson r correlation between the CES-D scores given by the SDB nurse and the psychiatric research assistant. For each patient's first pair of CES-D evaluations, the correlation was r = .76 (n = 24, p < .001). Inter-observer agreement, relative to the criterion score of 16, can also be seen from table 1. In 71 percent of all evaluations the two CES-D scores were in agreement on whether or not the patient was depressed. Since most of the agreements were in the null (not depressed) category, the Kappa coefficient was only .32 (p < .06). When the criterion cutoff score was changed to 20, the Kappa coefficient increased to .51 (p < .005)
indicating moderate agreement, because disagreements were reduced by almost half.

To measure internal consistency, the Spearman rho correlations between each item and the total test score were calculated on the CES-D data obtained from the SDB nurse. All but one of the correlations were significant (p < .05), ranging from 0.39 to 0.75. The correlation between CES-D item “I was depressed” and the total CES-D score was .72. The one item that was not significantly correlated with the total score was “I felt disliked”, rho = .32, p = .11.

The hypothesis that some items whose depression scores may be obscured by true physical inabilities (e.g. “I felt that everything I did was an effort”) would correlate less well with the total test score was not supported. The item-test correlations between the five items considered most likely to be affected by stroke (items 1,5,7,13,16) were not significantly lower than the item-test correlations for the other 15 items (Mann-Whitney U test, 1-tailed, n.s., mean ranks equaled .37-40. The same pattern was obtained for the set of first evaluations per patient as for the full data set, being approximately .10 higher for the latter. This pattern of correlations provided strong support for the construct validity of the CES-D in this context.

For the purpose of validating the CES-D, the most important relationship was between the CES-D and the Psychiatric Diagnosis (based on the DSM-III), and they correlated significantly, rho = .77, (p < .0001). The relationship between CES-D score and psychiatric diagnosis is shown on table 3. As the CES-D score rose, the DSM classification tended to move from “no depression” (CES-D ≤ 12) to “minor depression” (CES-D ≤ 24) to “major depression” (CES-D ≥ 25).

The ultimate test of the validity of the CES-D was to see how well it could distinguish between no depression and minor or major depression, (i.e. its sensitivity and specificity for depression). For the present sample, the estimated specificity of the CES-D was 100% (table 4). Whenever the CES-D score was equal to or greater than 16, the corresponding psychiatric diagnosis was depression. Thus the CES-D appeared to yield no false positives; i.e. in no case was a CES-D ≥ 16 indicative of a non-depressed patient. The estimated

<table>
<thead>
<tr>
<th>Psychiatric research assistant’s CES-D</th>
<th>SDB nurse’s CES-D</th>
<th>SDB nurse’s CES-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;16</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>≥16</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Psychiatric diagnosis, DSM-III†</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Zung depression scale</td>
<td>.65†</td>
<td></td>
</tr>
<tr>
<td>Hamilton depression test</td>
<td>.57†</td>
<td></td>
</tr>
<tr>
<td>Present state exam</td>
<td>.74‡</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other measures</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Social functioning exam</td>
<td>.09(ns)</td>
<td></td>
</tr>
<tr>
<td>Social ties checklist</td>
<td>−.04(ns)</td>
<td></td>
</tr>
<tr>
<td>John Hopkins functioning inventory</td>
<td>−.08(ns)</td>
<td></td>
</tr>
<tr>
<td>Mini mental state examination</td>
<td>.17(ns)</td>
<td></td>
</tr>
</tbody>
</table>

* Spearman’s rho, p < .0001.
† p < .002.
‡ p < .0001.

The Validity of CES-D for Stroke Patients

The CES-D correlated highly and significantly with all of the other measures of depression and did not correlate significantly with any of the other measures (table 2). The same pattern was obtained for the set of first evaluations per patient as for the full data set, being approximately .10 higher for the latter. This pattern of correlations provided strong support for the construct validity of the CES-D in this context.
sensitivity of the SDB CES-D at the 16 cutpoint was 73%. When the CES-D was less than 16, a very high proportion of the patients were not depressed (64%), but a number of depressed patients were missed (27%) false negative rate. The predictive value of a negative test result was 84% (16/19) for this sample, however the true predictive value of the CES-D will be a function of the prevalence of post-stroke depression.

Discussion

The present study has demonstrated that the CES-D, as administered by a research nurse to stroke patients, has good inter-rater reliability as well as concurrent, discriminant and construct validity, and is a useful screening tool for depression in this population.

The research nurse who administered the CES-D did not have extensive training in the evaluation or diagnosis of depression. She did, however, have prior experience in interviewing and evaluating stroke patients. Thus, although it is possible that the self-administration of this questionnaire by stroke patients or the administration by someone not experienced in giving patient questionnaires may not produce the high level of reliability found in this study, it is likely that this instrument can be a useful screening tool for the detection of depression in non-aphasic stroke patients.

Using a cut-off score of 16 in this sample the CES-D had 100% specificity and 73% sensitivity for the DSM-III diagnosis of either major or minor (dysthyemic) depression. This suggests that, if the CES-D were used as a screening instrument, any patient who scored above 16 would have a clinically important depressive disorder. Approximately 16% (3/19) of the patients who scored less than 16 on the CES-D, however, would also have a clinically significant depressive disorder, if the prevalence of post-stroke depression is about 40%. Thus, using a cut-off score of 16 on the CES-D, there would be some false negative patients, but few if any false positive patients.

In a previous publication, it has been demonstrated that post-stroke depressive disorders which meet the diagnostic criteria for either major or minor depression can be successfully treated with the tricyclic antidepressant nortriptyline.6 Untreated depressions have been demonstrated to be long lasting as well as severe with obvious significant emotional suffering. There are also a sizeable number of patients who have suicidal thought associated with severe depressive disorders. Thus, it is important that patients with clinically important depressive disorders be identified and evaluated for the possibility of treatment of their disorder. The CES-D Scale provides a relatively rapid and easy method of identifying patients with depressive disorders. It is hoped that this instrument will be an aid in the evaluation and recognition of commonly occurring depressive disorders in stroke patients.

A final note of qualification is in order. Because approximately 20% of the stroke patients also suffer from varying degrees of aphasia,24 the CES-D cannot be applied to a large segment of the target population. Furthermore, since aphasia is symptomatic of left hemisphere lesions, and the frequency of clinically significant depression is greater among patients with left frontal hemispheric lesions,5 it is likely that the prevalence of depression is greater among aphasics. A non-verbal, behavioral assessment screening measure for these patients still remains to be developed.

APPENDIX 1 The CES-D Scale

The interview below is given by the data bank nurse. The depression scale response cards are shown to the patient, who is asked, for each item, to pick the card "which best describes how you felt or behaved this past week." The nurse interviewer records the patient's response on this form.

Depression Scale

1. I was bothered by things that usually don't bother me. 0 1 2 3
2. I did not feel like eating; my appetite was poor. 0 1 2 3
3. I felt that I could not shake off the blues even with help from my family or friends. 0 1 2 3
4. I felt that I was just as good as other people. 0 1 2 3
5. I had trouble keeping my mind on what I was doing. 0 1 2 3
6. I felt depressed (blue or down). 0 1 2 3
7. I felt that everything I did was an effort. 0 1 2 3
8. I felt hopeful about the future. 0 1 2 3
9. I thought my life had been a failure. 0 1 2 3
10. I felt fearful. 0 1 2 3
11. My sleep was restless. 0 1 2 3
12. I was happy. 0 1 2 3
13. I talked less than usual. 0 1 2 3
14. I felt lonely. 0 1 2 3
15. People were unfriendly. 0 1 2 3
16. I enjoyed life. 0 1 2 3
17. I had crying spells. 0 1 2 3
18. I felt sad. 0 1 2 3
19. I felt that people disliked me. 0 1 2 3
20. I could not get "going" 0 1 2 3

CES-D Score

Items 4, 8, 12 and 16 have their values reversed before totaling.

Acknowledgments

The authors wish to thank Joshua Barwick and Anita Roth who assisted in the analysis of the data and preparation of the manuscript. We also acknowledge the helpful comments of Dr. M.B. Denckla of NINCDS.

References

Screening for depression in stroke patients: the reliability and validity of the Center for Epidemiologic Studies Depression Scale.
D Shinar, C R Gross, T R Price, M Banko, P L Bolduc and R G Robinson

Stroke. 1986;17:241-245
doi: 10.1161/01.STR.17.2.241

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/17/2/241