# Comparison of Six Depression Rating Scales in Geriatric Stroke Patients

Berit Agrell, MD, and Ove Dehlin, MD, PhD

We compared three self-rating scales (the Geriatric Depression Scale, the Zung Scale, and the Center for Epidemiologic Studies Depression Scale) with three examiner-rating scales (the Hamilton Rating Scale, the Comprehensive Psychopathological Rating Scale–Depression, and the Cornell Scale), to see which was best for 40 elderly (mean age 80 years) stroke patients, 17 of whom were depressed according to clinical examination. External validity and concurrent validity were good for all except the Cornell Scale. Reliability (internal consistency) showed that some items were not significantly correlated, which might be explained by our selection of the patients. The Geriatric Depression Scale, the Zung Scale, and the Comprehensive Psychopathological Rating Scale–Depression had the highest sensitivity, and the Zung Scale had the highest positive predictive value (93%). With regard to internal consistency, sensitivity, and predictive value, the best self-rating scales were the Geriatric Depression and the Zung scales and the best examiner-rating scale was the Comprehensive Psychopathological Rating Scale–Depression. (*Stroke* 1989;20:1190–1194)

After stroke, psychic complications, particularly depression, are of great importance with regard to rehabilitation and quality of life.[1,2] The prevalence of depression after stroke has been estimated to be 25–50%.[3,4] In a comparison of stroke patients and orthopedic patients with the same degree of physical handicap, depression was four times more common among the stroke patients.[5] Some studies have shown relations up to 1 year after stroke between the prevalence and severity of depression and the localization of a brain lesion in the frontal area of the left hemisphere.[1,6,7] Factors such as the degrees of physical handicap and cognitive impairment and the quality of social support available are related to the prevalence of poststroke depression, but such factors explain only a minor part of the variance.[6] Thus, it is probable that the physical handicap does not cause depression but that depression, when present, may accentuate the physical impairment and lead to poorer recovery and delayed rehabilitation.[1,8]

As depression after stroke is common, there is need for a valid and reliable rating scale for screening purposes. We compared three self-rating scales with three observation, or examiner-rating, scales to find which is best for geriatric stroke patients. A self-rating scale has an advantage in that it does not require extensive observation of the patient; a disadvantage, on the other hand, is that some elderly patients may have difficulty answering the questions and may require assistance in filling out the form. Some stroke patients also have cognitive impairment that can make using a self-rating scale difficult. Therefore, we also included a scale especially constructed for demented patients. The three self-rating scales we used were the Geriatric Depression Scale (GDS),[9] constructed especially for use in elderly subjects, the Zung Scale,[10] one of the most frequently used self-rating scales, and the Center for Epidemiologic Studies Depression Scale (CES-D).[11] Both the Zung Scale and the CES-D have been validated for stroke patients.[11,12] The three examiner-rating scales we used were the Hamilton Rating Scale (HRS),[13] which is well known and with which new scales are often validated; the Comprehensive Psychopathological Rating Scale–Depression (CPRS-D)[14]; and the Cornell Scale, designed especially for demented patients.[15] The HRS and the CPRS-D have both been used in studies of stroke patients.[1,8,12,16] For validity testing, we compared scores on the scales with a clinical examination (expressed as a global rating), with each other, and with certain patient characteristics.

## Subjects and Methods

Forty patients (18 men and 22 women) volunteered to take part in our study. Their mean age was 80 (range 61–93) years. All 40 patients had had a stroke 4 months to 2.5 years (mean 14 months)

before the study. The patients came from an outpatient day hospital, a geriatric rehabilitation clinic, or a nursing home. The Mini Mental State Examination (MMSE) was used as a cognitive test; the recommended cutoff point for this age group is 20.[17] This limit was not absolute, and we included patients even if they scored <20 if they could participate satisfactorily. We excluded patients with dysphasia or severe cognitive impairment since they were considered unable to answer questions or to fill out a self-rating scale. For two scales (the CES-D and the Cornell Scale), there were only 39 and 38 patients, respectively.

The Barthel Index was used to evaluate activities of daily living (ADL) function.[18] The brain lesion was localized with computed tomography or (in some cases) on clinical grounds only. The frequency of visits from relatives or friends was used as a measure of social network and was classified as 1/day, <1/day to 1/week, <1/week to 1/2 weeks, and <1/2 weeks. Data regarding earlier stroke or psychiatric disease and treatment with antidepressants or the analgesics acetaminophen, propoxyphene, or nonsteroidal anti-inflammatory drugs were obtained from the patient, from the medical records, or from the nursing staff.

The clinical examination or psychiatric interview was used as the independent variable to which we related the scales. The results of the clinical examination were expressed as a global rating (no illness, minimal depression that does not interfere with social functioning, unmistakable but moderately expressed depression, and serious or disabling depression).[14] The clinical examination was performed and all scales were applied for each patient on 2–3 successive days by the same investigator. Accuracy of scoring was expressed as very low, doubtful, good, or very good.[14]

The study was approved by the Ethics Committee of the University of Lund. All patients were given written and oral information about the study, and all patients gave verbal consent before entering the study.

One-way analysis of variance (ANOVA) was used to compare groups of patients having low ($\leq20$) and higher (>20) MMSE scores with accuracy of scoring. One-way ANOVA was also used to test differences in the scores on each scale with respect to MMSE score, time since last stroke, Barthel Index score, localization of the brain lesion, frequency of visits, history of stroke or psychiatric disease, and treatment with antidepressants or analgesics. Discriminant validity was assessed by correlating the scores on each scale with the MMSE score, the Barthel Index score, frequency of visits, history of stroke or psychiatric disease, and treatment with antidepressants or analgesics.

External validity was calculated as the Pearson product–moment correlation between the scores on the scales and the global ratings; concurrent validity was calculated as the Pearson product–moment correlation between the scores on the six scales. The Pearson product–moment correlation was also calculated for the global ratings and the scores on all six scales, the MMSE score, the Barthel Index score, and age. Sensitivity, specificity, and predictive value of a positive or negative depression test were calculated according to standard formulas. Global ratings were used as true positives and true negatives. A two-tailed probability value of $\leq0.05$ was considered significant.

Reliability of each scale (internal consistency) was tested by calculating Spearman rank correlation coefficients between the score on each item and the sum of the scores for each scale. The corresponding probability values for these coefficients were calculated according to Conover.[19] Internal consistency was also calculated as Cronbach's $\alpha$.[20]

## Results

Most patients could fill out the questionnaires without help, but it was necessary to explain individual items to some patients. In the HRS, information for Item 14 (loss of libido) was often lacking or inadequate and thus was rated as 0, as recommended by Hamilton.[13] In the Zung Scale, Item 6 (I still enjoy sex) was replaced by the mean score for all items in this scale since nearly all patients were unable to answer this item satisfactorily.

Patient characteristics are shown in Table 1. No patient had aspirin for prevention of stroke recurrence. Seventeen patients (43%) were depressed as assessed by the global rating (Table 1).

Seven patients had a low MMSE score. Higher MMSE scores correlated significantly with greater accuracy of scoring ($p<0.01$). The MMSE score was significantly correlated with age and Barthel Index score; higher age and more pronounced handicap were related to impaired cognitive function ($p<0.05$ and $p<0.01$, respectively). There were no significant differences among the scales with respect to MMSE scores, time from last stroke, frequency of visits, history of stroke or psychiatric disease, or treatment with antidepressants. The mean scores for all six scales were higher for patients with right- than for patients with left-hemisphere lesions, but the difference was not significant ($p=0.1$). Low Barthel Index ($\leq50$) score was significantly associated with high scores on the Cornell Scale ($p<0.005$), and patients treated with analgesics scored significantly higher on the HRS ($p<0.05$). Discriminant validity was thus satisfactory for the GDS, the Zung Scale, the CES-D, and the CPRS-D since no correlation was found between any of them and patient characteristics.

External and concurrent validity are shown in Table 2. With the exception of the Cornell Scale, validity was good. Sensitivity, specificity, and predictive value at the recommended cutoff points are shown in Table 3. The Zung Scale had the highest and the Cornell Scale the lowest positive predictive value.

TABLE 1.    **Characteristics of 40 Geriatric Stroke Patients**

| Characteristic | Women ($n=22$) | Men ($n=18$) | Total No. | % |
|---|---|---|---|---|
| MMSE score | | | | |
| Mean | 23 | 24 | — | — |
| Range | 14–30 | 14–30 | — | — |
| Barthel Index score | | | | |
| Mean | 33 | 46 | — | — |
| Range | 0–100 | 10–90 | — | — |
| Brain lesion | | | | |
| Right hemisphere | 16 | 13 | 29 | 73 |
| Left hemisphere | 5 | 4 | 9 | 22 |
| Other | 1 | 1 | 2 | 5 |
| Antidepressants | 4 | 2 | 6 | 15 |
| Analgesics | 12 | 11 | 23 | 58 |
| Global rating | | | | |
| No illness | 13 | 10 | 23 | 57 |
| Minimal depression, does not interfere with social functioning | 5 | 3 | 8 | 20 |
| Unmistakable but moderately expressed depression | 3 | 5 | 8 | 20 |
| Serious or disabling depression | 1 | 0 | 1 | 3 |

MMSE, Mini Mental State Examination used as cognitive test; recommended cutoff point for persons in this age group is 20. Data are number of patients unless noted otherwise.

Internal consistency showed that some items in a scale were not significantly correlated with the sum of the scores for that scale. There were four such items in the GDS: Items 1 (Are you basically satisfied with your life?), 2 (Have you dropped many of your activities and interests?), 12 (Do you prefer to stay at home, rather than going out and doing new things?), and 14 (Do you feel you have more problems with memory than most?). In the Zung Scale there were two, Items 4 (I have trouble sleeping at night) and 9 (My heart beats faster than usual). In the CES-D there were three such items: 10 (I felt fearful), 15 (people were unfriendly), and 19 (I felt that people disliked me). In the HRS there were five: items 5 (insomnia, middle), 8 (retardation), 16 (loss of insight), 19 (derealization and depersonalization), and 20 (paranoid symptoms). In the CPRS-D, all items correlated significantly and eight of 10 items were very highly correlated ($p<0.001$). Cronbach's $\alpha$ was 0.90 for the GDS, 0.83 for the Zung Scale, 0.64 for the CES-D, 0.81 for the HRS, and 0.89 for the CPRS-D. Internal consistency was not calculated for the Cornell Scale as it was found to have an unsatisfactory external validity. The receiver operating characteristic curves of the scales are compared in

TABLE 2.    **External and Concurrent Validity of Six Depression Rating Scales for Geriatric Stroke Patients**

| | Scale | | | | | |
|---|---|---|---|---|---|---|
| | Self-rating | | | Examiner-rating | | |
| | GDS | Zung | CES-D | HRS | CPRS-D | Cornell |
| External validity | | | | | | |
| Global rating | 0.75 | 0.72 | 0.73 | 0.65 | 0.86 | 0.23* |
| Concurrent validity | | | | | | |
| Self-rating scales | | | | | | |
| GDS | | | | | | |
| Zung | 0.88 | | | | | |
| CES-D | 0.82 | 0.81 | | | | |
| Examiner-rating scales | | | | | | |
| HRS | 0.77 | 0.70 | 0.74 | | | |
| CPRS-D | 0.86 | 0.82 | 0.83 | 0.87 | | |
| Cornell | 0.37† | 0.32‡ | 0.31§ | 0.26¶ | 0.35‖ | |

Data are Pearson product–moment correlations for 40 patients (except $n=39$ for CES-D and $n=38$ for Cornell). GDS, Geriatric Depression Scale; CES-D, Center for Epidemiologic Studies Depression Scale; HRS, Hamilton Rating Scale; CPRS-D, Comprehensive Psychopathological Rating Scale–Depression.
*†‡§¶‖$p<0.17$, 0.02, 0.05, 0.06, 0.1, 0.03, respectively; all others $p<0.001$ different from 0.

TABLE 3. Sensitivity, Specificity and Predictive Value of Positive or Negative Test at Recommended Cutoff Points for Six Depression Rating Scales for Geriatric Stroke Patients

| Scale | n | Recommended cutoff (points) | Sensitivity (%) | Specificity (%) | Predictive value Positive (%) | Negative (%) |
|---|---|---|---|---|---|---|
| Self-rating | | | | | | |
| GDS | 40 | 10 | 88 | 64 | 58 | 88 |
| Zung | 40 | 45 | 76 | 96 | 93 | 84 |
| CES-D | 39 | 20 | 56 | 91 | 82 | 75 |
| Examiner-rating | | | | | | |
| HRS | 40 | 10 | 71 | 87 | 60 | 80 |
| CPRS-D | 40 | 3.0 | 88 | 70 | 68 | 89 |
| Cornell | 38 | 8 | 44 | 64 | 47 | 61 |

GDS, Geriatric Depression Scale; CES-D, Center for Epidemiologic Studies Depression Scale; HRS, Hamilton Rating Scale; CPRS-D, Comprehensive Psychopathological Rating Scale–Depression.

Figure 1.[21] The CPRS-D and the Zung Scale had the best sensitivity and specificity.

## Discussion

A self-rating scale should not be too long or too time-consuming, especially for a depressed patient who also might be psychomotor retarded. Most of the patients were able to answer all items, with two exceptions. One was Item 6 in the Zung Scale (I still enjoy sex), which must be considered inappropriate for this patient population. The other was the item about sex in the HRS, Item 14 (loss of libido), which was also nearly always impossible to get adequate information about.

Accuracy of scoring correlated significantly with cognitive function measured by MMSE scores. This is in agreement with the general clinical experience that it is more difficult to evaluate a demented patient than a nondemented one.[15]

All scales showed good discriminant validity, the only exceptions being the Cornell Scale for patients



FIGURE 1. *Receiver operating characteristic curves for six depression rating scales for geriatric stroke patients. GDS, Geriatric Depression Scale; CES-D, Center for Epidemiologic Studies Depression Scale; HAM, Hamilton Rating Scale; CPRS-D, Comprehensive Psychopathological Rating Scale–Depression.*

with decreased mobility ($p<0.005$) (Barthel Index Score ≤50) and the HRS for patients treated with analgesics ($p<0.05$). Discriminant validity is important in scoring stroke patients and elderly patients as they often have more than one disease, impaired cognition, and decreased ADL function.

External validity was highest for the CPRS-D, although all scales except the Cornell Scale showed significant validity. The Cornell Scale was constructed for use in demented patients, whereas severe dementia was an exclusion criteria in our population, which might explain that scale's low validity in our study. Also, in the Cornell Scale no score should be given if symptoms result from physical disability or illness, and most patients in our study had rather pronounced somatic illnesses. The method of administration is also unique for the Cornell Scale; it is based on two interviews, one with the caregiver or nurse and one with the patient, whereas in the two other examiner-rating scales, information is obtained from a structured or semi-structured interview. The Cornell Scale, with its special design, was not appropriate in this patient population with no or only mild to moderate cognitive impairment.

Concurrent validity was high for all scales except the Cornell Scale. In other studies, significant correlations of similar magnitudes have been found (e.g., between the GDS and the Zung Scale, 0.82[9]; between the GDS and the HRS, 0.82[9]; between the Zung Scale and the CES-D, 0.65[11]; between the Zung Scale and the HRS, 0.79 and 0.80[7,22]; between the CES-D and the HRS, 0.57[11]; and between the HRS and the CPRS-D, 0.89.[23] The assertion that the concurrent validity between the HRS and the self-rating scales is low could not be verified in this study.[22]

The prevalence of poststroke depression in our study was 17 of 40, 43%, comparable to findings in other studies.[12,24] Most depressed patients had minor or dysthymic depressions, and only one patient had severe depression according to the global rating (Table 1). Only six of the 17 depressed patients were treated
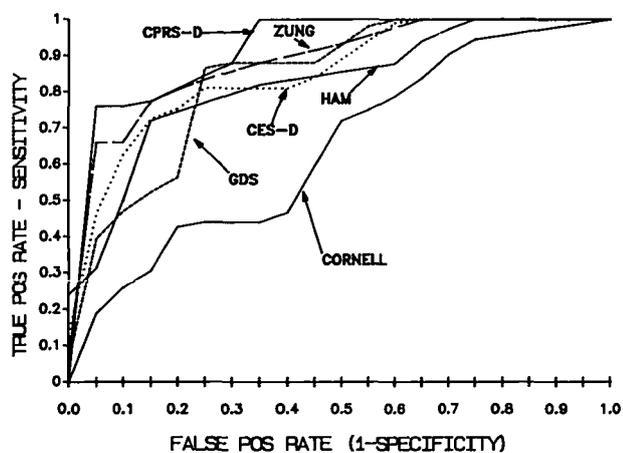
with antidepressants, which shows that the diagnosis was often missed. It is important to keep the prevalence of depression in mind when discussing predictive value, as this varies with prevalence. All tests and evaluations of each patient were done within 2–3 days. This could reduce the patients' interest in doing the tests and thus the internal consistency. On the other hand, the risk for mood changes was small. The population we studied is in certain respects selected; cognitive impairment was only mild or moderate, patients with dysphasia sufficient to impair communication were excluded, and 73% had right-hemisphere brain lesions. This limits the generalizability of our results. In addition to handicaps due to their stroke, most patients also had compensated chronic diseases, which is common at this high age. The high rate of use of analgesics reflects the patients' high age and diseases but possibly also the local prescribing tradition.

The cutoff point of a scale should maximize both sensitivity and specificity and can be shifted according to the investigator's wish to increase sensitivity or specificity.[9] A low cut off point increases the number of depressed patients diagnosed but increases the number of false-positives. As shown in Figure 1, all but the Cornell Scale had acceptable sensitivity as well as good specificity.

The GDS and the CPRS-D showed the best internal consistency according to Chronbach's $\alpha$. The specific characteristics of the population under study may explain the differences we found in internal consistency. Classic depressive symptoms appear in three domains: affective symptoms, cognitive symptoms, and somatic symptoms such as disturbances in energy level, appetite, sleep, libido, and elimination. In geriatric depression, somatic symptoms (especially fatigue, constipation, and insomnia) are prominent, but their diagnostic specificity is reduced by the fact that many nondepressed elderly persons report the same symptoms.[9] In the Zung Scale, the two items that did not show significant internal consistency were somatic symptoms. In the HRS, one of the five items not significantly correlated with the sum of the scores was a somatic symptom. In the CES-D there were three items that did not correlate, and one of these (I felt that people disliked me) also failed to correlate significantly in an earlier study of stroke patients.[11] Another explanation for the fact that sleep-related items were not correlated may be that many patients were treated with hypnotics. Depression-induced sleep disturbances could have consequently been masked.

Which scale, then, is best suited to screening for depression in elderly stroke patients? The GDS, the Zung Scale, and the CPRS-D all have high sensitivity, from 88% to 76%, at the recommended cut off points. The Zung Scale has the highest positive predictive value, 93%, and, along with the GDS and the CPRS-D, the greatest internal consistency. Our results from this patient population indicate that the GDS and the Zung Scale are the best self-rating scales and that the CPRS-D is the best examiner-rating scale.

## References

1. Robinson RG, Starr LB, Lipsey JR, Rao K, Price TR: A two year longitudinal study of poststroke mood disorders: Dynamic changes in associated variables over the first six months of follow-up. *Stroke* 1984;15:510–516
2. Ahlsjö B, Britton M, Murray V, Theorell T: Disablement and quality of life after stroke. *Stroke* 1984;15:886–890
3. Wade DT, Legh-Smith J, Hewer RA: Depressed mood after stroke. *Br J Psychiatry* 1987;151:200–205
4. Robinson RG, Lipsey JR, Price T: Diagnosis and clinical management of post-stroke depression. *Psychosomatics* 1985; 26:769–778
5. Folstein MF, Mailberger R, McHugh PR: Mood disorder as a specific complication of stroke. *J Neurol Neurosurg Psychiatry* 1977;40:1018–1020
6. Robinson RG, Chait RM: Emotional correlates of structural brain injury with particular emphasis on post-stroke mood disorders. *Crit Rev Clin Neurbiol* 1985;1:285–319
7. Robinson RG, Starr LB, Kubos KL, Price TR: A two year longitudinal study of poststroke mood disorder: Findings during the initial evaluation. *Stroke* 1983;14:736–741
8. Parikh RM, Lipsey JR, Robinson RG, Price TR: Two-year longitudinal study of poststroke mood disorders: Dynamic changes in correlates of depression at one and two years. *Stroke* 1987;18:579–584
9. Brink TL, Yesavage JA, Lum O, Heersema PH, Adey M, Rose TL: Screening tests for geriatric depression. *Clin Gerontol* 1982;1:37–43
10. Zung WWK: A self rating depression scale. *Arch Gen Psychiatry* 1965;12:63–70
11. Shinar D, Gross CR, Price TR, Banko M, Bolduc PL, Robinson RG: Screening for depression in stroke patients: The reliability and validity of the Center for Epidemiologic Studies Depression Scale. *Stroke* 1986;17:241–245
12. Robinson RG, Szetela B: Mood change following left hemispheric brain injury. *Ann Neurol* 1981;9:447–453
13. Hamilton M: Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;6:278–298
14. Åsberg M, Montgomery SA, Perris C, Schalling D, Sedvall G: A comprehensive psychopathological rating scale. *Acta Psychiatr Scand [Suppl]* 1978;271:5–27
15. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA: Cornell Scale for depression in dementia. *Biol Psychiatry* 1988;23:271–284
16. Viitanen M: Long-term effects of stroke (dissertation). Umeå University, 1987
17. Folstein MF, Folstein SE, McHugh PR: "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–198
18. Mahoney FJ, Barthel DW: Functional evaluation: Barthel Index. *Md State J* 1965;14:61–65
19. Conover WJ: *Practical Nonparametric Statistics*, ed 2. New York, John Wiley & Sons, Inc, 1980
20. Cronbach LJ: Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334
21. Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman SH, Klerman GL: Performance of screening and diagnostic tests. *Arch Gen Psychiatry* 1987;44:550–555
22. Hamilton M: Comparative value of rating scales. *Br J Clin Pharmacol* 1976;1(suppl):58–60
23. Montgomery SA, Åsberg M: A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134:382–389
24. Feibel JH, Springer CJ: Depression and failure to resume social activities after stroke. *Arch Phys Med* 1982;63:276–277

# Stroke

American Heart Association | American Stroke Association®

## Comparison of six depression rating scales in geriatric stroke patients.
B Agrell and O Dehlin

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/20/9/1190