J. psychiat. Res., Vol. 17, No. 1, pp. 37–49, 1983.
Printed in Great Britain.

0022-3956/83/010037-13 $03.00/0

DEVELOPMENT AND VALIDATION OF A GERIATRIC DEPRESSION SCREENING SCALE: A PRELIMINARY REPORT

JEROME A. YESAVAGE, T. L. BRINK

Department of Psychiatry and Behavioral Sciences, Stanford University of Medicine, Stanford, CA 94305, U.S.A.

TERENCE L. ROSE

Veteran's Administration Medical Center, Palo Alto, CA 94304, U.S.A.

OWEN LUM

Geriatric Treatment Team, Santa Clara County Mental Health

VIRGINIA HUANG, MICHAEL ADEY and VON OTTO LEIRER

Veteran's Administration Medical Center, Palo Alto, CA 94304, U.S.A.

(*Received* 25 *January* 1982; *revised* 28 *June* 1982)

Summary—A new Geriatric Depression Scale (GDS) designed specifically for rating depression in the elderly was tested for reliability and validity and compared with the Hamilton Rating Scale for Depression (HRS-D) and the Zung Self-Rating Depression Scale (SDS). In constructing the GDS a 100-item questionnaire was administered to normal and severely depressed subjects. The 30 questions most highly correlated with the total scores were then selected and readministered to new groups of elderly subjects. These subjects were classified as normal, mildly depressed or severely depressed on the basis of Research Diagnostic Criteria (RDC) for depression. The GDS, HRS-D and SDS were all found to be internally consistent measures, and each of the scales was correlated with the subject's number of RDC symptoms. However, the GDS and the HRS-D were significantly better correlated with RDC symptoms than was the SDS. The authors suggest that the GDS represents a reliable and valid self-rating depression screening scale for elderly populations.

INTRODUCTION

MOST EXISTING depression rating scales have been developed and validated in younger populations and their applicability with older persons has not yet been demonstrated. The scale described in this article was specifically designed to measure depression in the aged, primarily as a screening instrument, and validated within this population.

MEASURING DEPRESSION IN THE AGED

The need for a geriatric depression scale is obvious. Between 5 and 20% of the 20 million aged Americans are estimated to be depressed (GURLAND, 1976). Although one could apply existing general psychiatric depression scales to this population, the aged present unique problems for clinicians and researchers interested in the study and treatment of depression (SALZMAN and SHADER, 1978).

A major problem is the confusion of dementia with depression in the elderly. The syndrome of "pseudodementia", with psychomotor retardation and passive refusal to

respond appropriately to cognitive tests is depression mistaken for dementia (WELLS, 1979; JARVIK, 1976). Depression in the elderly often is accompanied by subjective experiences of memory loss and cognitive impairment (KAHN *et al.*, 1975), symptoms seen less frequently in the young.

Conversely, somatic symptoms which are usually a key to diagnosis of depression in the young are less useful in the elderly. For instance, sleep disturbances are a common symptom of endogenous depression; but such disturbances are also common in the nondepressed elderly (COLEMAN *et al.*, 1981), while rare in younger persons not suffering from depression. A host of other examples include the normal decline of sexual function, constipation, and the aches and pains associated with arthritis in the aged.

The high prevalence of somatic complaints among the elderly and their unique cognitive complaints present both a problem and an opportunity in screening for depression in the elderly. The problem is that most existing scales are heavily loaded toward measuring the somatic symptoms of depression. Although somatic complaints are clearly part of major depressive disorders, this will not necessarily be the case in milder forms of depression. Moreover, to the extent one is interested in screening for depression rather than formal diagnosis or description, discrimination between depressed and nondepressed persons or between different degrees of depression would seem to be the primary concern. For this reason it may be necessary to weight somatic symptoms of depression less heavily than psychological symptoms having greater discriminative power. On the other hand, the unique cognitive complaints of the elderly may present an opportunity to devise screening instruments with enhanced discriminative power in the elderly.

Another problem in the assessment of geriatric depression and other disorders experienced by the aged is that the elderly are typically more resistant to psychiatric evaluation than younger patients (SALZMAN and SHADER, 1978; WELLS, 1979). Consequently, one needs to design the items comprising a scale to fit this population; questions appropriate for use with the young may not be appropriate for the old. For example, questions about sexuality often make the elderly defensive, and yet they are included on many existing scales. Other questions may pose problems of patient acceptance as well as leading to problems of interpretation (BLUMENTHAL, 1975). For example, questions about suicidal intent, whether life is worth living, or whether one is hopeful about the future obviously have different meaning in those reaching the end of their lifespan. Of course these problems of patient resistance and unique interpretation can probably be dealt with adequately if an experienced interviewer administers the depression scale, and the scale is designed to elicit more open-ended responses from the patient in an atmosphere fostering good rapport. However, in designing a self-rating depression scale for the aged, these issues need to be adequately addressed in the scale's initial development.

It is also essential to provide a simple, easily understood format in the development of a geriatric depression scale. Several of the self-rating scales presently available may be too difficult for the elderly to complete by themselves. For example, Zung's (1965) self-rating scale for depression uses a four-point scale that is likely to be more confusing than a yes/no format, because it involves a greater number of choices and subtle discriminations that must be made by the person.

The scale reported here was designed to avoid most of these problems associated with

the measurement of geriatric depression by developing the scale with the aged in mind and by selecting items for the scale based on their performance within this population. Questions that proved to have inadequate power to discriminate depressed from non-depressed elderly were not incorporated into the scale while uniquely discriminating items that might not be as useful with younger groups were included. Furthermore, a yes/no format was used in order to make the scale a simple one that could be used in nearly all instances as a self-rating scale and one that would be acceptable to patient and physician alike.

EXISTING DEPRESSION SCALES

There are numerous depression rating scales currently available. These have been subject to several reviews (CARROLL *et al.*, 1973; KOCHANSKY, 1979; MCNAIR, 1979; HEDLUND and VIEWEG, 1979) and include: Hamilton Rating Scale for Depression (HRS-D), Zung Self-Rating Depression Scale (SDS), Beck Depression Inventory, Phenomena of Depression Scale, Grading Scale for Depressive Reactions, Psychiatric Judgment of Depression Scale, NIMH Collaborative Depression GDS, SAD-GLAD, Verdum Depression Rating Scale, CES-D, SCL-90 Profile of Mood States, and the MMPI Depression Scale.

These scales represent a mixture of observer-rated and self-rating scales. In some cases the same scale may also have been designed to function in either manner or an observer-rated measure of depression has been adapted for use as a self-rating scale. CARROLL *et al.*'s (1973) adaptation of the HRS-D represents an example of the latter approach. A problem with adapting a scale from one format to the other, however, is that questions which may have been acceptable to respondents in an interview setting where good rapport is established by the interviewer may no longer be accepted by the respondent when the same questions are asked using a self-rating scale. We have found this to be the case, for example, with Carroll *et al.*'s scale; mildly depressed subjects dislike the disease-oriented questions and have difficulty with questions which assume they are in a hospital setting.

However, the primary problem with these depression scales is that they were not originally designed for use with the elderly and rarely have they been properly validated in this population. There are some exceptions. There has been an attempt to validate the SDS in the aged, but the ability of the SDS to discriminate nondepressed from depressed elderly was found to be limited (ZUNG and GREEN, 1973). Zung suggested using a classification criterion of 40 for depression; although this would correctly identify 88% of depressives, it leads to the false identification of normal elderly as being depressed in 44% of the cases. Other comprehensive reviews suggest that there are still no better criteria that would reduce the number of false positives associated with the SDS (CARROLL *et al.*, 1973; CARROLL, 1978). Thus, although this represents the best validation efforts in this population to date, the SDS still has limitations as a geriatric depression screening device.

Despite the virtual absence of studies aimed at validating existing depression scales within elderly populations, these scales may prove to be useful even though they were not originally designed with the aged in mind. For this reason two of the existing scales were included in the present research. Their inclusion also was desirable so that comparisons between the GDS with currently existing measures could be made. The present research

was not aimed, however, at demonstrating the superiority of the proposed scale over those currently available. Indeed, the enormous number of existing scales would make this a tremendous undertaking. Rather, existing scales were included in order to provide additional information about the convergent validity of the GDS and to enable tentative norms for the GDS to be compared to those for other, more extensively researched measures.

The first of these was the Hamilton Rating Scale for Depression or HRS-D (HAMILTON, 1960). It was intended to be a measure of treatment outcome rather than as a screening device. Unlike the GDS it is designed to be completed by an experienced observer after a 30 min clinical interview which assays most phenomena associated with "endogenous" depression, e.g. insomnia, decreased libido, loss of appetite (LYERLY, 1978). The HRS-D is probably the most widely accepted clinical interview for depression. It has been shown to be a rapidly learned and reliable measure (HAMILTON, 1967) capable of distinguishing between different degrees of depression (CARROLL *et al.*, 1973; BIGGS *et al.*, 1978; KNESEVICH *et al.*, 1977) and to be one of the few scales available that is also useful as a diagnostic instrument (SCHNURR *et al.*, 1976).

The other scale included in the present research was the Zung Self-Rating Scale for Depression or SDS (ZUNG, 1965). It was administered because of its popularity and the availability of norms for elderly subjects. The SDS has been found to be internally consistent with split-half reliability coefficients in the range of 0.73–0.79. However, validity coefficients have shown greater variability across studies; correlations with the HRS-D have ranged from 0.22 to 0.95 (HEDLUND and VIEWEG, 1979). Although quite widely used among clinicians and researchers working psychiatry, the SDS has recently come under criticism as both a research measure and clinical screening device (CARROLL *et al.*, 1973).

Two studies were conducted in the process of developing and validating the Geriatric Depression Scale (GDS). In the first study, a large pool of items were constructed and then tested for the extent to which they appeared to measure depression in the aged. In the second subject, a subset of these items were selected, readministered to a new sample of subjects, and validated against an independent criterion of depression. The latter study also provided a basis for comparing properties of the GDS to other existing measures of depression due to the inclusion of the SDS and HRS-D. These studies will be discussed in turn. Finally, after describing the results of these studies, a number of recent investigations aimed at demonstrating the performance of the GDS in more specific elderly populations will be discussed.

STUDY ONE: ITEM SELECTION

Methods

A team of clinicians and researchers involved in geriatric psychiatry selected 100 questions believed to have potential for distinguishing elderly depressives from normals. In choosing these questions care was taken to include material covering a wide variety of topics relevant to depression, such as somatic complaints, cognitive complaints, motivation, future/past orientation, self-image, losses, agitation, obsessive traits, and mood itself. A yes/no format was chosen for ease of administration since our experience

with the SDS indicated that a range of possibilities often confused elderly patients. Questions also were phrased in a format that would not alarm patients or make them overly defensive. We thought that these features of the scale would maximize its use as a self-rating instrument of depression in the elderly.

After selecting these items for inclusion in the questionnaire, it was administered in its self-rating form to 47 subjects. The subjects were either normal elderly living in the community with no complaints of depression and no history of mental illness, or subjects hospitalized for depression. Both male and female patients were included from a number of hospitals in Santa Clara County California. All subjects were over 55 years old.

Results

Data analysis was based on the rationale that the 100 item scale should have *prima facia* validity for depression and that those items which correlated best with the total score would be most likely to measure depression. The 30 items (Table 1) correlated highest and most significantly with the total score were chosen for inclusion in the GDS. The median correlation among these items was 0.675 (range = 0.47–0.83). For the 100-item, the median correlation was 0.51 (range = −0.07 to 0.83).

TABLE 1. GERIATRIC DEPRESSION SCALE

Choose the best answer for how you felt over the past week

1.	Are you basically satisfied with your life?	yes / no
2.	Have you dropped many of your activities and interests?	yes / no
3.	Do you feel that your life is empty?	yes / no
4.	Do you often get bored?	yes / no
5.	Are you hopeful about the future?	yes / no
6.	Are you bothered by thoughts you can't get out of your head?	yes / no
7.	Are you in good spirits most of the time?	yes / no
8.	Are you afraid that something bad is going to happen to you?	yes / no
9.	Do you feel happy most of the time?	yes / no
10.	Do you often feel helpless?	yes / no
11.	Do you often get restless and fidgety?	yes / no
12.	Do you prefer to stay at home, rather than going out and doing new things?	yes / no
13.	Do you frequently worry about the future?	yes / no
14.	Do you feel you have more problems with memory than most?	yes / no
15.	Do you think it is wonderful to be alive now?	yes / no
16.	Do you often feel downhearted and blue?	yes / no
17.	Do you feel pretty worthless the way you are now?	yes / no
18.	Do you worry a lot about the past?	yes / no
19.	Do you find life very exciting?	yes / no
20.	Is it hard for you to get started on new projects?	yes / no
21.	Do you feel full of energy?	yes / no
22.	Do you feel that your situation is hopeless?	yes / no
23.	Do you think that most people are better off than you are?	yes / no
24.	Do you frequently get upset over little things?	yes / no
25.	Do you frequently feel like crying?	yes / no
26.	Do you have trouble concentrating?	yes / no
27.	Do you enjoy getting up in the morning?	yes / no
28.	Do you prefer to avoid social gatherings?	yes / no
29.	Is it easy for you to make decisions?	yes / no
30.	Is your mind as clear as it used to be?	yes / no

Although twelve of the 100 original items assessed somatic complaints (e.g. sleep disturbance, anorexia, weight loss, cardiac or gastrointestinal symptoms), none of these were among the 30 items which correlated strongest with the total score. The median correlation between the somatic items and the total score was 0.33 (range = 0.02–0.45). Thus, these items were excluded from the final scale, because they did not meet the purely empirical criterion adopted as a basis for an item's inclusion.

Of the 30 questions selected for inclusion in the GDS, 20 indicated the presence of depression when answered positively while ten others (Nos 1, 5, 7, 9, 15, 19, 21, 27, 29 and 30) indicated depression when answered negatively. The questions were arranged in a 30 item, one-page format and ordered so as to maximize patient acceptance of the questionnaire. Having arrived at a final version of the GDS, a validation study was implemented.

STUDY TWO: VALIDATION

Method

Two groups of geriatric subjects were chosen for the validation phase. The first of these ($n = 40$) consisted of normal elderly persons recruited at local senior centers and housing projects. These subjects had no histories of mental illness and were functioning well in the community. The second group ($n = 60$) consisted of subjects under treatment for depression. These subjects were both inpatients and outpatients, male and female, and in Veterans Administration, county and private treatment settings.

The subjects under treatment were further differentiated into mild and severe depression groups. The frequently used criteria of outpatient vs inpatient groups was not used because in some settings, such as the county mental health service, many severe depressives were outpatients while in other settings, such as the Veterans Administration, many mild depressives were inpatients. Instead, it was decided to divide our group of clinically depressed subjects into mild and severe groups on the basis of whether or not they met Research Diagnostic Criteria (RDC) for a major affective disorder (depressed) (SPITZER *et al.*, 1978). These criteria, elicited during a clinical interview, involve eight symptoms: weight loss, sleep difficulty, loss of energy, psychomotor retardation, loss of interest or pleasure in usual activities, feelings of self-reproach or guilt, complaints of diminished ability to concentrate and recurrent thoughts of death or suicide. Five are required to make the diagnosis. Using these criteria it was possible to separate the depressives into a "mild" group ($n = 26$), having an average of 3.4 RDC criteria symptoms, and a "severe" group ($n = 34$) with an average of 5.9 RDC criteria symptoms. These two groups then became our second and third subject groups, respectively.

The subjects in all groups were given a clinical interview lasting 30–60 min which involved a rating of the HRS-D and the administration of the two self-rating scales, the SDS and our GDS. The interviews were conducted by trained observers, the authors. Interrater reliability on the HRS-D was 0.9. For those subjects who were unable to complete the self-rating scales without assistance, the examiner read the questions orally, elicited answers from the subject, and recorded his or her responses. The order in which the scales were administered was randomly determined for each subject.

Results

Internal consistency and reliability

Four measures of internal consistency were computed for each of the three depression scales. These included: (1) the median correlation between the individual items comprising a scale and the corrected-item total score (total score minus score on the particular time involved); (2) the average intercorrelation among the scale's individual items; (3) CHRONBACH's (1951) alpha coefficient; and (4) the split-half reliability coefficient. Each of these measures or indices of interanl consistency provides a basis for judging the extent to which the scale's items all measure the same underlying construct. In addition to computing these various indices of internal consistency, test–retest data are reported for the GDS. These data provide information regarding the reliability, i.e. stability, of GDS scores over time.

The results of the internal consistency analyses are displayed in Table 2. Each of the indices computed for the depression scales are discussed in turn below.

TABLE 2. COMPUTED INDICES OF INTERNAL CONSISTENCY FOR THE GDS, SDS AND HRS-D

Index	Scale		
	GDS	SDS	HRS-D
Median correlation with total score	0.56	0.44	0.56
Mean interitem correlation	0.36	0.25	0.34
Alpha coefficient	0.94	0.87	0.90
Split-half reliability	0.94	0.81	0.82

Correlation with total score. The median correlation between the items of the GDS and the corrected-item total scores was 0.56 (range = 0.32–0.83), suggesting that all of the items on this scale do, in fact, measure a common latent variable. The comparable values for the SDS and HRS-D were 0.44 (range = 0.24–0.71) and 0.56 (range = 0.16–0.81), respectively. Based on these data it would appear that the GDS, HRS-D and SDS are all internally consistent measures.

Inter-item correlations. The mean intercorrelation among items from the GDS was 0.36; the computed values for the SDS and HRS-D were 0.25 and 0.34, respectively. These values are in a range necessary for a high degree of internal consistency for each scale as a whole, as confirmed by the analyses which follow.

Alpha coefficient. CHRONBACH's (1951) alpha coefficient was utilized in order to provide an overall measure of the internal consistency of the GDS. The computed value of the alpha coefficient was 0.94, suggesting a high degree of internal consistency for the GDS. Computed values of the alpha coefficient for the SDS and HRS-D were 0.87 and 0.90, respectively.

Split-half reliability. An alternative index of internal consistency is the split-half reliability coefficient. This measure is typically derived by splitting a scale into two equivalent forms, calculating their intercorrelation, and then estimating the reliability of the composite scale using the Spearman–Brown formula (NUNNALLY, 1967). Employing this procedure, the reliability coefficients for the GDS, SDS, and HRS-D were found to be 0.94, 0.81, and 0.82, respectively. These values are reported in order to allow comparisons with previous research.

Test–retest reliability. Test–retest reliability was calculated for the GDS by having 20 subjects complete the questionnaire twice, one week apart. A correlation of 0.85 was obtained ($p < 0.001$), suggesting that, at least within the time frame considered here, scores on the GDS reflect stable individual differences.

Validity

The primary test of the validity of the GDS as a measure of depression was provided by the classification of subjects as normal (i.e. nondepressed), mildly depressed, or severely depressed on the basis of RDC for major affective disorder. If both this classification variable and the GDS are valid indices of depression, one would expect normal subjects to receive the lowest GDS scores whereas severely depressed subjects should score the highest on this measure. As a test of this hypothesis, an analysis of variance was conducted in which the classification variable served as a between-subjects factor while the subjects' total scores on the GDS served as the dependent measure. Similar analyses were also performed on the SDS and HRS-D. The results of these analyses provided evidence for each of the scales' validity. In each analysis the main effect for the classification variable was highly significant [GDS: F (2, 97) = 99.48, $p < 0.001$; SDS: $F(2, 97) = 44.75$, $p < 0.001$; HRS-D: $F(2. 97) = 110.63$, $p < 0.001$], and as seen in Table 3, in each case the means were ordered as predicted. t-Tests conducted between each pair of means within the same row of this table showed that subjects classified as normal scored significantly lower on each of the scales compared to the mildly and severely depressed subjects while the severely depressed group scored higher than each of the other two groups (all $p < 0.001$). These findings, then, provide evidence

TABLE 3. MEANS AND STANDARD DEVIATIONS FOR THE GDS, SDS, AND HRS AS A FUNCTION OF SUBJECT CLASSIFICATION

Scale	Group Normal	Mildly depressed	Severely depressed	Total sample
GDS	5.75 (4.34)	15.05 (6.50)	22.85 (5.07)	13.98 (9.02)
SDS	34.31 (6.66)	44.15 (11.39)	52.79 (7.51)	43.15 (11.53)
HRS	5.43 (4.98)	13.35 (5.98)	25.42 (6.45)	14.29 (10.35)

*Standard deviations appear in parentheses.

for the validity of the GDS as a measure of depression as well as validating the SDS and HRS-D.

Given previous findings indicating that the SDS (ZUNG, 1965; HEDLUND and VIEWEG, 1979) and HRS-D (CARROLL *et al.*, 1973; HAMILTON, 1960, 1967; BIGGS *et al.*, 1978; KNESEVICH *et al.*, 1977) are valid measures of depression, positive correlations between these measures and the GDS would provide evidence for the scales' convergent validity. The obtained correlation between the GDS and the SDS was found to be 0.84 while a correlation of 0.83 was found between the GDS and the HRS-D. The correlation between the SDS and the HRS-D was 0.80. All of these correlations were statistically reliable at or beyond the 0.001 level.

These analyses provided additional evidence of the validity of each of these depression scales. However, given the criticism that the SDS often may not adequately distinguish between different levels of depressive symptomatology (CARROLL *et al.*, 1973) a comparison was also made across the three scales to determine the relative strength with which each one was related to the RDC. The correlation of each of the depression scales with the classification variable derived from these criteria was computed, and then, following FERGUSON (1971), the magnitude of each correlation was compared to the other two. The obtained correlations between the classification variable and the GDS, SDS, and HRS-D were 0.82, 0.69, and 0.83, respectively. All of these represented statistically reliable correlations (all p's < 0.001). However, comparing each of these correlations to the others showed that, whereas those associated with the GDS and the HRS-D did not differ significantly from each other, t (97) < 1, both of these were significantly greater in magnitude than that associated with the SDS [GDS vs SDS: t (97) $= 3.83$, $p < 0.001$; HRS-D vs SDS: t (97) $= 3.85$, $p < 0.01$]. It thus appears that, compared to the other two measures, the SDS discriminates less effectively between the normal, mildly depressed, and severely depressed subjects.

DISCUSSION

These results provide evidence that the GDS is a reliable and valid measure of geriatric depression. A high degree of internal consistency was found for the scale, and total scores on the GDS were reliable over a one-week interval. Evidence for the validity of the scale came from a comparison of the mean scores associated with subjects classified as normal, mildly depressed, or severely depressed based on RDC criteria for depression; the three groups' means were reliably different and ordered as one would expect given their differing RDC scores.

The primary purpose for constructing the GDS was to provide a reliable screening test for depression in elderly populations that would be simple to administer and not require the time or skills of a trained interviewer. The fact that the GDS was found to discriminate between groups of normal, mildly depressed, and severely depressed subjects is encouraging in this regard. However, one would ultimately desire information on the percentage of individuals correctly and incorrectly classified using particular scores on this measure. This can be accomplished by computing indices of *sensitivity* and *specificity* for the measure, where in this case sensitivity refers to the number of depressed persons correctly classified as depressed based on a particular criterion and where specificity refers

to the number of nondepressed persons correctly classified as such. Sensitivity is lowered to the extent depressed persons are missed using a criterion and classified incorrectly as nondepressed whereas specificity declines to the extent nondepressed persons are incorrectly labelled as suffering from depression.

Sensitivity and specificity of the GDS was examined in a recent study conducted by our research group (BRINK *et al.*, 1981). It was found that among elderly persons drawn from the same centers as those used in the present study, a cut-off score of 11 on the GDS yielded a 84% sensitivity rate and a 95% specificity rate. A more stringent cut-off score of 14 yields a slightly lower, 80%, sensitivity rate, but results in the complete absence of nondepressed persons being incorrectly classified as depressed, i.e. a 100% specificity rate. Based on these findings BRINK *et al.* (1981) suggested that a score of 0–10 be viewed as within the normal range while 11 or greater being a possible indicator of depression. Criteria for the SDS and HRS-D were also offered; these were scores of 46 and 11, respectively. A score of 46 on the SDS achieves 80% sensitivity and 85% specificity whereas a score of 11 on the HRS-D achieves 86% sensitivity and 80% specificity. The three scales, however, are best compared by holding either sensitivity or specificity constant. With specificity held constant at 80%, the sensitivities of the GDS, SDS and HRS-D were found to be 90, 82, and 86%, respectively.

A geriatric depression scale should not only be applicable for screening depression in the physically healthy elderly but should also be useful with the physically ill, and cognitively impaired. There is some evidence that the GDS may fulfill this criterion. Using data from a study by GALLAGHER *et al.* (1981), we found that the GDS differentiated depressed from nondepressed elderly in a sample of subjects who all suffered from physical illness. These subjects were elderly arthritics who had been given the GDS after having been classified as either depressed or nondepressed based on a comprehensive clinical interview. Comparing the GDS scores of these two groups of arthritics it was found that the mean score of the depressed subjects (13.1, S.D. = 7.14) was indeed significantly higher than that of the nondepressed subjects (5.10, S.D. = 4.21), $t\,(47) = 4.94$, $p < 0.001$. These data, then, provide evidence that the validity of the GDS is not limited to elderly subjects who are physically healthy.

In another recent study the GDS was found to differentiate depressed from non-depressed elderly undergoing cognitive treatment for senile dementia. These subjects were classified as demented by criteria of FOLSTEIN *et al.*'s (1975) Mini-Mental Status Exam. It was found that those subjects categorized as depressed by a therapist blind to GDS scores received a mean score of 14.72 (S.D. = 6.13) on the GDS vs a mean of only 7.49 (S.D. = 4.26) for nondepressed subjects, $t\,(41) = 4.4$, $p < 0.001$. Although the results of this study should only be viewed as suggestive since the number of subjects was small ($n = 43$), this study provides preliminary evidence that the GDS is a valid measure of depression with demented, as well as normal, elderly subjects.

However, despite evidence for each of the three scales' validity, they did not appear to perform equally well with respect to the task of differentiating between various RDC defined degrees of depression. Because the GDS and HRS-D were correlated with the number of RDC symptoms each subject had to a significantly greater extent than the SDS, one could argue that, among the two self-rating scales, the GDS appears to provide a

more sensitive screening instrument. Although the SDS was found to correlate more poorly with the RDC than either the GDS or HRS-D, differences in the content and format of the three scales should be considered in making this comparison. It is important to recognize, first of all, the similarity between the three scales and the criterion, the RDC. The HRS-D would be expected to be more strongly related to the RDC, and the group classification variable, than the other two scales simply because the RDC are heavily represented on the HRS-D. Thus, the GDS and SDS are at a disadvantage in the analyses undertaken in the present study, because they do not measure all of the symptoms comprising the RDC while measuring others (e.g. diurnal symptom variation) which are not reflected in these criteria. Moreover, the poorer performance of the SDS may have been due partly to the fact that the RDC measure the severity of depression while the SDS measures the frequency of symptoms, and the two may not correspond closely (CARROLL *et al.*, 1973).

The RDC were chosen as the basis for classifying the level of depression in subjects because of a consensus among researchers that it appears to capture the essential aspects of depressive disorders. Given its wide acceptance, and the lack of a better set of criteria, the failure of a scale to correlate well with the RDC probably reflects more upon the scale in question than the RDC. However, despite the differences in content between the RDC and the GDS, the GDS total score was found to still correlate as strongly with the number of RDC symptoms as the HRS-D whose content corresponds more closely with these criteria. Thus, emphasizing the subjective aspects of depression rather than the somatic and behavior aspects does not seem to have detracted from the validity of the GDS as it may have in the case of the SDS. Despite the differences in content betwen the GDS and HRS-D, the former scale did nevertheless appear to be as valid as the HRS-D in the present research. This finding is somewhat surprising given the absence of somatic symptoms on the former and reliance upon them in the latter. This may be explained in part by the fact that both scales assay mood dysphoria and other psychological symptoms of depression, which seem to best discriminate between the depressed and nondepressed aged.

The issue of how well somatic items measure depression in the elderly and discriminate the depressed from nondepressed is one which deserves further attention. In the first study of the present series, the somatic items' median correlation with the total score was only 0.33, compared to 0.68 for the selected questions. A similar pattern emerged for both the SDS and HRS-D in the second study. On the SDS the items most highly correlated with the corrected-item total score were those concerned with the subjective, psychological aspects of depression while the items most poorly correlated with the total score were those dealing with the somatic aspects of depression. The four lowest correlations were those measuring constipation, decreased libido, appetite decrease, and somatic anxiety while the four highest correlations were those measuring personal devaluation, emptiness, depressed mood, and dissatisfaction. Nearly identical findings have been obtained by STEUER *et al.* (1980). They found total scores on the SDS to be most highly correlated with those items measuring dissatisfaction, depressed mood, emptiness and personal devaluation whereas the lowest correlations occurred with those items measuring constipation, somatic anxiety, decreased libido, and agitation. Moreover, they found

further evidence that the somatic items of the SDS may measure depression more poorly than subjective states in elderly patients by computing four sets of factor scores labelled well being, depressed mood, optimism, and somatic symptoms. Not only was the somatic factor correlated the least strongly with the SDS total score, but this factor was the only one found to be significantly correlated with physician health ratings. Thus, this study demonstrates how, even among individuals screened for serious illness, the poorer health of the aged may undermine the power of somatic symptoms to detect depression.

Of course it is possible that the SDS simply does not contain good measures of those somatic symptoms accompanying depressive illness. But this interpretation does not explain the findings in which other measures of depression have been utilized. For example, although less marked than the results found with the SDS, a similar pattern of results was found in the present research when the items from the HRS-D were correlated with the total score: the somatic items generally correlated less strongly than the items measuring loss of interests, depressed mood and anxiety. Similarly, DESSONVILLE *et al.* (1981) using the Schedule for Affective Disorders and Schizophrenia (SADS) have found that, even though the somatic aspects of depression differentiated depressed from nondepressed elderly, the mean differences between the two groups were smaller on the somatic items than those measuring the subjective states of depression.

Clearly more research is needed on the expression of depression within elderly subjects. The fact that the subjects in the present research were all relatively healthy, as were subjects in these additional studies, may have preserved the discriminability of some somatic questions. It remains to be determined whether the somatic items on these scales adequately measure depression in elderly persons who are less healthy. The GDS appears to avoid many of these problems by focusing on the psychological aspects of depression. This is not meant to imply that somatic symptoms should not be measured in cases of depressive illness. Such symptoms need to be assessed when one is concerned with formal diagnosis or when there is the desire to examine changes in the expression of depressive illness. However, when screening is the goal, discrimination between levels of depression is of primary important and somatic questions may be less powerful in this regard than items chosen empirically for their ability to differentiate the nondepressed from the depressed.

Finally, it is important to distinguish instruments to be used for screening, diagnosis and assessment of change. As the above data indicate, all three may find use as screening instruments, even if this was not their original intent. None, however, is a diagnostic tool. Positive results on any of the three scales on screening should be followed up by a clinical interview if significant levels of depressive symptomatology are found and treatment is being considered. On the other hand, the HRS-D has also been shown to be quite sensitive to changes in the level of symptomatology over time (KNESEVICH *et al.*, 1977), and thus, may serve well, as it was originally intended, as a means of gauging changes in the severity of depression. The use of the SDS in outcome research is more controversial (CARROLL *et al.*, 1973; CARROLL, 1978). It remains to be determined if the GDS may be useful for measuring changes in the severity of depression following treatment.

In conclusion, though not a substitute for observer-rated scales or indepth diagnostic interviews, and not yet shown to be treatment sensitive, the GDS appears to be a promising and simple screening instrument which may find other applications through further research.

Acknowledgement—This research was supported by the Medical Research Service of the Veteran's Administration.

REFERENCES

BIGGS, J. T., WYLIE, L. T. and ZIEGLER, V. E. (1978) Validity of the Zung Self-Rating Depression Scale. *Br. J. Psychiat.* **132,** 381–385.

BLUMENTAL, M. D. (1975) Measuring depressive symptomatology in a general population. *Archs gen. Psychiat.* **32,** 971–978.

BRINK, T. A., YESAVAGE, J. A., LUM, O., HEERSEMA, P., ADEY, M. and ROSE, T. L. (1982) Screening tests for geriatric depression. *Clin. Gerontologist* **1,** 37–44.

CARROLL, B. J. (1978) Validity of the Zung Self-Rating Scale (letter to the editor). *Br. J. Psychiat.* **133,** 379.

CARROLL, B. J., FIELDING, J. M. and BLASHKI, T. G. (1973) Depression rating scales: a critical review. *Archs gen. Psychiat.* **28,** 361–366.

CHRONBACH, L. J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* **16,** 297–334.

COLEMAN, R. M., MILES, L. E., GUILLEMINAULT, C., ZARCONE, V. P., VAN DER HOED, J. and DEMENT, W. C. (1981) Sleep-wake disorders in the elderly: a polysomnographic analysis. *J. Am. Geriat. Soc.* **29,** 289–296.

DESSONVILLE, C. L., GALLAGHER, D., THOMPSON, L., FINNELL, K. and LEWINSOHN, P. (1982) Relationship of age, health status, and depressive symptoms in normal and depressed older adults. *Essence* **5,** 99–117.

FERGUSON, G. A. (1971) *Statistical Analysis in Psychology and Education,* 3rd edn., pp. 171–172. McGraw-Hill, New York.

FOLSTEIN, M. F., FOLSTEIN, S. E. and MCHUGH, P. R. (1975) Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J. psychiat. Res.* **12,** 189–198.

GALLAGHER, D., SLIFE, B. and YESAVAGE, J. A. (1983) Impact of physical health status on scores of older adults on the Hamilton rating scale for depression. Under Editorial review.

GURLAND, B. J. (1976) The comparative frequency of depression in various adult age groups. *J. Geront.* **31,** 283–292.

HAMILTON, M. (1960) A rating scale for depression. *J. Neurol. Neurosurg. Psychiat.* **23,** 56–62.

HAMILTON, M. (1967) Development of a rating scale for primary depressive illness. *Br. J. Soc. Clin. Psychol.* **6,** 278–296.

HEDLUND, J. L. and VIEWEG, B. W. (1979) The Zung Self-Rating Depression Scale: a comprehensive review. *J. opl Psychiat.* **10,** 51–64.

JARVIK, L. F. (1976) Aging and depression: some unanswered questions. *J. Geront.* **31,** 324–326.

KAHN, R. L., ZARIT, S. H., HILBERT, N. M. and NIEDEREHE, G. (1975) Memory complaint and impairment in the aged: the effect of depression and altered brain function. *Archs gen. Psychiat.* **32,** 1569–1573.

KNESEVICH, J. W., BIGGS, J. T., CLAYTON, P. J. and ZIEGLER, V. E. (1977) Validity of the Hamilton Rating Scale for Depression. *Br. J. Psychiat.* **131,** 49–52.

KOCHANSKY, G. E. (1979) Psychiatric rating scales for assessing psychopathology in the elderly: a critical review. In *Psychiatric Symptoms and Cognitive Loss in the Elderly: Evaluation of Assessment Techniques* (Edited by RASKIN, A. and JARVIK, L. F.), pp. 125–156. Wiley, New York.

LYERLY, S. B. (1978) *Handbook of Psychiatric Rating Scales,* 2nd edn. National Institute of Mental Health, Rockville, Maryland.

MCNAIR, D. M. (1979) Self-rating scales for assessing psychopathology in the elderly. In *Psychiatric Symptoms and Cognitive Loss in the Elderly: Evaluation and Assessment Techniques* (Edited by RASKIN, A. and JARVIK, L. F.), pp. 157–167. Wiley, New York.

NUNNALLY, J. (1967) *Psychometric Theory,* p. 194. McGraw-Hill, New York.

SALZMAN, C. and SHADER, R. I. (1978) Depression in the elderly: relationship between depression, psychologic defense mechanisms and physical illness. *J. Am. Geriat. Soc.* **26,** 253–259.

SCHNURR, R., HOAKEN, P. C. S. and JARRETT, F. J. (1976) Comparison of depression inventories in a clinical population. *Can. psychiat. Ass. J.* **21,** 473–476.

SPITZER, R. L., ENDICOTT, J. and ROBINS, E. (1978) Research Diagnostic Criteria: rationale and reliability. *Archs gen. Psychiat.* **35,** 773–782.

STEUER, J., BANK, L., OLSEN, E. J. and JARVIK, L. F. (1980) Depression, physical health and somatic complaints in the elderly: a study of the Zung Self-Rating Depression Scale. *J. Geront.* **35,** 683–688.

WELLS, C. E. (1979) Pseudodementia. *Am. J. Psychiat.* **136,** 895–900.

ZUNG, W. W. K. (1965) A self-rating depression scale. *Archs gen. Psychiat.* **12,** 63–70.

ZUNG, W. W. K. and GREEN, R. L., JR. (1973) Detection of affective orders in the aged. In *Psychopharmacology and Aging* (Edited by EISDORFER, C. and FANN, W. E.), pp. 213–223. Plenum Press, New York.