The Geriatric Depression Scale: A Review of Its Development and Utility

Ignacio Montorio and María Izal

ABSTRACT. This article reviews the significance of the Geriatric Depression Scale (GDS) to practitioners and researchers in clinical gerontology, more than 10 years after the scale was introduced to the scientific community. This report summarizes findings from the most relevant validation studies in which this self-report for assessing depression in elderly people has been tested. Included is discussion of the use of the GDS with specific populations (elderly medical inpatients, nursing home residents, and dementia populations), with description of the scale's psychometric properties and its utility when used with them. This article also provides data on the use of the GDS from more recent studies, including additional information on psychometric properties, influence of source bias, and the international dissemination of the GDS. We conclude that the GDS is a relevant self-report for the assessment of depression in the elderly, given its advantage over other self-reports that are not as easily administered to this age group, its utility in the detection of depression, and its adequate psychometric properties. However, the GDS does not maintain its validity in demented populations because it fails to identify depression in persons with mild to moderate dementia. Finally, some suggestions for future research are made.

Depression in old age is a topic that has produced a great amount of literature that has substantially contributed to the improvement of the assessment process. The ambiguity of the definition of depression in old age, its differentiation from other disorders (e.g., dementia), and the role that other related factors play in the diagnosis (e.g., medication intake) have been identified as common problems that should be investigated. Difficulties regarding assessment procedures are as important as those previously mentioned. There is a fair amount of literature about the influence of source bias on self-report. However, the self-report (meaning source of information: self rather than informant) is more traditional at measuring depression than other assessment procedures, because the characteristics of depression are largely subjective (e.g., dysphoric mood) (Rehm, 1988).

The use of self-report for assessing depression in the elderly shares the aforementioned characteristics with the assessment of depression in other adults. However, specific problems have been identified in the use of self-report...
I. Montorio and M. Izal

for assessing depression in old age. For example, one of the most frequent sources of error when using self-report in the elderly is social desirability (respondent's wish to present himself or herself in a favorable way; Klassen et al., 1975). Some aspects of the content of general depression self-reports could also raise problems, such as somatic complaints or health or death concerns, when applied to the elderly.

GDS DEVELOPMENT AND VALIDITY

In the assessment of depression in old age, the Geriatric Depression Scale (GDS) (Brink et al., 1982; Yesavage et al., 1983) is currently one of the most used depression self-reports. Because the GDS was created for the elderly, its items were based on characteristics of depression in the elderly (Coleman et al., 1981; Jarvik, 1976; Wells, 1979). The GDS was developed and validated in two studies (Brink et al., 1982; Yesavage et al., 1983). In the study of Brink and colleagues (1982), the initial conception of the GDS used the rational criteria of researchers and clinicians involved in geriatric psychiatry and gerontology. These experts selected 100 items with yes/no answers that had been shown to be useful in distinguishing elderly depressed subjects from elderly normal subjects. An empirical selection was then made of 30 items that showed higher correlations with the total score on the 100-item scale when applied to a sample of 100 elderly volunteers living in the community. None of the final 30 items was somatic (although 12 of the 100 original items had been), thus avoiding one of the problems with self-reports assessing depression in the elderly, namely the confusion of somatic symptoms with physical disturbances that are common in old age.

In a study by Yesavage and colleagues (1983), the 30-item scale was subsequently validated against scores from two assessment instruments of depression. This study included measures of depression from the Zung Self-Rating Scale for Depression (SDS; Zung, 1965) and the Hamilton Depression Rating Scale (HAMD; Hamilton, 1967) (a version of the Hamilton rating scale converted into a self-report), to provide a basis for comparing properties of the GDS with these measures. This study was carried out with two samples, one of 60 depressed elderly subjects (patients complaining of depression) and the other of 40 elderly subjects not affected by depression (normal subjects without histories of mental illness). Those with depression were, in addition, divided into "severe" and "mild" cases using the Research Diagnostic Criteria (RDC; Spitzer et al., 1978). Statistical analyses were then conducted to examine the concurrent validity of the three depression scales used. The correlations between the classification criteria ("no depression," "mild depression," and "severe depression") and each of the scales, GDS, SDS, and HAMD, were \( r = .82, r = .69, \) and \( r = .83, \) respectively, all of them being statistically significant (all \( p < .001 \)).

In regard to the sensitivity and specificity of the GDS (range of possible scores = 0 to 30), among samples of subjects chosen from the same centers as
those used in the study of Yesavage and colleagues (1983), a score of 11 or more correctly classified 84% of depressed elderly (sensitivity) and 95% of those not affected by depression (specificity). A more restrictive cutoff score of 14 gave an 80% sensitivity rate and a 100% specificity rate. In light of these data, Yesavage and colleagues (1983) suggested that a score of 0 to 10 should be considered normal and 11 or more as a possible indicator of depression. Data from a subsequent study have confirmed the appropriateness of similar cutoffs for the GDS (Hickie & Snowdon, 1987). In this study, each subject was interviewed to determine the presence or absence of a major depressive episode using criteria of the Diagnostic and Statistical Manual of Mental Disorders (3rd ed.; DSM-III; American Psychiatric Association, 1980). A score of 14 gave an 88% sensitivity rate and 84% specificity rate, and a score of 15 gave an 88% sensitivity rate and a 100% specificity rate. However, these results must be interpreted with caution because the subjects were drawn from different populations (day centers, nursing homes, psychiatry outpatients, and a general psychiatric ward). Finally, other interesting data come from a study by Gallagher, which revealed that the GDS differentiated depressed and nondepressed elderly people, even though all the subjects in the sample had some physical illness (Gallagher, Slife, & Yesavage, 1986, cited in Sheikh & Yesavage, 1986).

TYPES OF POPULATIONS

The validity of the GDS has been analyzed mostly with regard to elderly persons living independently in the community. The diagnostic accuracy of the GDS when used in other living arrangements must also be considered. First, differences among the results from various studies of elderly medical inpatients have been found. Thus, in the first study analyzing the GDS when used with this population (Rapp et al., 1988), results indicated moderate diagnostic accuracy. When a GDS cutoff score of 14 as denoting depression was employed, the sensitivity rate was 65% and the specificity rate was 93%. In a subsequent study with elderly medical inpatients, more satisfactory results on the diagnostic accuracy of the GDS were obtained: Using 14 as the cutoff score, the sensitivity rate was 92% and the specificity rate was 89% (Koenig et al., 1988).

Second, the psychometric properties of the GDS when it is used with the elderly in nursing homes are not as satisfactory as with the community elderly. Although the results for the reliability of the GDS with institutionalized elderly people are consistent with those found in the original research (alpha coefficient = .99; test-retest reliability = .94; Lesher, 1986) and the validity of the GDS is supported by its convergence with the Beck Depression Inventory (BDI; \( r_{xy} = .78 \)) (Kiernan et al., 1986), the sensitivity and specificity of the scale when it is used with institutionalized elderly people are not consistent, particularly for those with intermediate levels of depression (Lesher, 1986; Parmelee et al., 1989). When 11 was used as a cutoff score, the sensitivity rate was 69% for elderly people with depressive symptoms or subclinical depression and 100% for elderly persons with major depression, and the specificity rate was 74%. 
When a cutoff score of 14 was used, the sensitivity rate for those with depressive features ranged from 55% to 100% for those with major depression, and the specificity rate was 81% (Lesher, 1986).

Third, the use of the GDS with elderly persons who have dementia has serious drawbacks. The first is in the administration of the GDS, because only elderly people with levels of 0 and 1 (representing cognitively intact and mildly demented individuals, respectively) on the Clinical Dementia Rating (Berg, 1984) are able to complete the GDS (Burke et al., 1989). Moreover, when the GDS is used with elderly persons who have mild and moderate levels of cognitive impairment, the diagnostic accuracy of the scale is low (the sensitivity rate is 25% and the specificity rate is 75%; Kafonek et al., 1989). Finally, the GDS may not be useful with dementia patients who are unaware of their cognitive deficit, because those dementia patients who disavow memory loss also tend to deny depressive symptoms on the GDS (Feher et al., 1992).

IMPORTANT FINDINGS

In addition to the aforementioned studies, most of which are original studies validating the use of the GDS among elderly persons with physical or cognitive impairments, other studies exist that yield interesting data on the GDS, mainly in terms of its validity and reliability. The following findings, although not exhaustive, are what we consider the most important results from these studies:

1. A 5-minute positive-theme reminiscence prior to administration of the GDS does not affect the test-retest reliability of the scale; the GDS thus does not appear to measure passing moods (Brink et al., 1985).

2. The correlations between the GDS and the BDI are high: Both measures are sensitive to treatment, so that the BDI and GDS may be assumed to evaluate the same overall construct of depression (Kiernan et al., 1986; Scogin, 1987). Likewise, there are data that confirm the convergent validity between the GDS and the Depression Adjective Checklists (DACL) (Izal & Montorio, 1993; Lubin, 1967) since a high correlation score between them has been found ($r_{xy} = .86, p < .001$).

3. Variations in reference group instructions that may be given for completion of the GDS do not significantly alter results (Dunn & Sacco, 1988).

4. A factorial analysis of the GDS generated five factors constituting the scale, which explain 42% of the total variance (Sheikh et al., 1991). The first factor, sad mood, reflects persistent thoughts of sadness; the second factor, lack of energy, includes cognitive complaints that are translated into difficulties in concentration and a lack of initiative; the third, positive mood, is related to positive affect and a positive worldview; the fourth, agitation, reflects different aspects of anxiety; and the fifth, social withdrawal, is associated with passivity and the avoidance of social situations. This factor structure provides clinicians with measures that are more descriptive than a simple total score. Thus, it may be a useful way of interpreting GDS scores because it characterizes the patient’s subjective experience of depression (Sheikh et al., 1991).
5. The correlation between the GDS and the Depression Symptoms Checklist (DSC, a self-report that incorporates the symptoms of depression from the DSM-III; \( r = 0.82 \)) was statistically significant \( (p < 0.01) \) and greater than that between the GDS and Zung's SDS \( (r = 0.59) \), or than that between the DSC and the SDS \( (r = 0.57) \) (Dunn & Sacco, 1988).

6. The GDS improves on Zung's SDS in that it provides a higher rate of completion of the scale \( (p < 0.5) \), so that the mean percentage of GDS items answered is 88%, whereas the mean percentage of SDS items answered is 79% (Dunn & Sacco, 1988).

7. Research with elderly populations often includes control or comparison groups consisting of people under 65, making it desirable to test the reliability of the GDS for adults under 65. A recent study on internal consistency (Cronbach's alpha coefficient) of the GDS with groups aged 30 to 39, 40 to 49, 50 to 59, and 60 to 69, sampled both in the community and in the university, showed high values (over 0.80) except for the university group aged 30 to 39 (Rule et al., 1989).

8. No relation has been found between the tendency to respond to items on the basis of social desirability, as measured by the Marlowe-Crowne Social Desirability Scale (M-CSDS), and the score attained with the GDS (Cappeliez, 1989).

9. The GDS has so far been translated and adapted into 12 languages, including Japanese (Niino et al., 1991), Portuguese (Zilenovski, 1991), Italian (Ferrario et al., 1990), and Spanish (González Felipe, 1988; González Felipe & Szurek Soler, 1990; Izal & Montorio, 1993; Perlado, 1987).

CONCLUSIONS

The use of self-reports for assessing depression in the elderly, as in other age groups, is widespread. However, it poses some problems that should be taken into account (Gallagher et al., 1980; Kane & Kane, 1981; Klassen et al., 1975; Lawton et al., 1980; Montorio, 1990; Zarit et al., 1985). To resolve these problems, self-reports for the elderly should be modified or specially designed with the following formal characteristics: They should be short, easily understood, and appropriate in terms of the size of letters in the items and in terms of the elderly person's level of education; they should include relevant age-related items; and they should provide normative data on the elderly population (McNair, 1979). Furthermore, the main sources of error found in relation to the use of self-reports among the elderly are relevance, social desirability, inhibition of response, anxiety, and understanding (Montorio, 1990, 1994).

Community-Dwelling Elderly

If clinicians keep in mind the above parameters, along with the classical indicators of reliability and validity, the GDS is a suitable self-report for the detection of depression among the elderly living in the community. In this
population, the GDS shows good capacity for discriminating between depressed and nondepressed elderly people, even in samples including elderly persons with various physical complaints. The GDS’s relevance is further seen in its capacity for measuring the same overall construct of depression as the most universal and widely studied scale for the assessment of depression in adults, the BDI (Beck et al., 1961). Moreover, as already pointed out, there is no tendency to respond to items on the basis of judgments of desirability of item content, so that, in short, the GDS is not influenced by the respondent’s social desirability (the tendency to present himself or herself in a favorable way). Also, the lack of inclusion of elements such as sexual orientation, the simplicity of the response procedure, and the GDS’s brief average length enhance understanding, diminish the anxiety generated by the administration of a psychological test, and produce a high completion rate (Pfeiffer, 1987; Zarit et al., 1985).

**Persons With Dementia**

However, although the GDS shows adequate diagnostic accuracy in elderly persons living in the community, it does not show the same accuracy for elderly persons in institutions. Therefore, the GDS should be used cautiously as a screening instrument in a population in which dementia is prevalent or in persons known to have dementia, because the GDS is not sensitive in detecting depression in demented subjects. Use of the GDS in dementia patients has been questioned because these patients, due to memory impairment, may not accurately recall their affective status for the past week as requested by the GDS instructions (Burke et al., 1989). This inaccuracy of recall confirms what had previously been stated, based on clinical experience, about the lack of validity of the GDS in dementia patients (Brink, 1984). The GDS has been found to have moderate to high diagnostic accuracy in differentiating depressed elderly medical inpatients from normal subjects (Koenig et al., 1988; Rapp et al., 1988).

On the other hand, the GDS has generated some controversy over not including items of a somatic nature. Somatic items could increase the risk of misdiagnosis of depression among the elderly, according to findings by Zung (1967). Examination of SDS items showed that somatic items contributed greatly to the total SDS score in the elderly, so it was concluded that somatic items could have a different significance for the elderly compared with other adults (Blumenthal, 1975; Gallagher et al., 1978). Some common physical illnesses in old age (such as neurologic, endocrine, or arthritic diseases) could have the same somatic symptoms as depression (for instance, endocrine disorders cause apathy, reduction of activity, and loss of energy). Furthermore, medication side effects could be confused with symptoms of depression, especially when the patient is prescribed neuroleptics, tranquilizers, or drugs for hypertension, Parkinson’s disease, or cancer (Salzman & Shader, 1978). In general terms, the more confusing symptoms are insomnia, weight loss, and absence of energy; in addition, the confusion between physical illness and depression is greater when physical illness is more serious (Dessonville et al.,
The GDS does not include items considered to measure somatic complaints of depression, therefore preventing misdiagnoses of depression due to confusion between somatic symptoms caused by depression and those produced by illness or medication intake. Nevertheless, further research is needed to determine whether some somatic complaints could be used to discriminate between depression and physical problems. The inclusion of such items in self-reports for depression like the GDS would increase the diagnostic accuracy of the reports.

Directions for Research

Several directions for future research on the GDS can be suggested. First, additional research on the GDS factor structure that may validate the underlying structure initially proposed for the scale is necessary (Sheikh et al., 1991). Second, further testing of the GDS’s sensitivity to treatments for depression in several settings (community, nursing homes, psychiatric wards) is needed. Third, because self-reports have shown some shortcomings in detecting depression among the elderly with dementia, the GDS should be administered and interpreted jointly with other assessment instruments, as recently recommended (Alexopoulos et al., 1988). Finally, although the GDS has been translated into several languages, no adequate validity studies have been carried out on all of these translations. Such research is needed not only to confirm the psychometric properties of the GDS when used in other languages, but also to initiate cross-cultural studies of the GDS.

More than 10 years after its publication, the GDS is an important contribution to gerontological research and clinical practice, and a suitable and promising tool for the assessment of depression in the elderly.

REFERENCES


Review of the GDS


Offprints. Requests for offprints should be directed to Ignacio Montorio, PhD, Department of Biological and Health Psychology, Autonomous University of Madrid, Madrid 28049, Spain.

---

**SUCCESSFUL GRANT WRITING**

*Strategies for Health and Human Service Professionals*

Laura N. Gitlin, PhD and Kevin J. Lyons, PhD

This book guides the reader through the language and basic components of grantmanship. It illustrates how to to develop ideas for funding, write the sections of a proposal, organize different types of project structures, and finally, to understand the review process.

Each chapter describes a specific aspect of grantmanship and suggests innovative strategies to implement the information that is presented. The appendices contain helpful materials, such as a list of key acronyms, examples of timelines and sample budget sheets. The strategies in this volume are beneficial to individuals and departments in academic, clinical, or community-based settings.

1996 235pp hard 0-8261-9260-2 $32.95 (outside US $37.50)

536 Broadway, New York, NY 10012-3955 • (212) 431-4370 • Fax (212) 941-7842