British Journal of Psychiatry (1994), **164**, 522–527

Semi-structured Depression Scale Sensitive to Change with Treatment for Use in the Elderly

A. V. RAVINDRAN, K. WELBURN and J. R. M. COPELAND

The construction of a semi-structured interview depression scale that is sensitive to change for use in the elderly is described. Depression items from a well validated diagnostic instrument, the Geriatric Mental State Schedule (GMSS), were used as the core items in the development of the instrument. Improvement in depression in 80 elderly patients was independently assessed with two standard rating scales for depression, the Hamilton Rating Scale for Depression and the Beck Depression Inventory, and by an independent clinician's judgement before and after standard antidepressant treatment. Depression items that were sensitive to change were retained from the core items to form the new instrument. Results indicate that this scale is reliable and valid, shows better correlation with both the clinician's and the patient's judgement of improvement than the standard instruments, and is sparing of the rater's time.

Depression is said to be the most common psychiatric diagnosis in the elderly (Post, 1982), and more and more of these patients are being treated with antidepressants. These clinical trials continue to use instruments devised for and validated in the younger population, such as the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1967), in spite of well documented differences in symptoms, treatment response, and residual symptoms in the elderly (Zung & Green, 1972; Post, 1982; Koenig *et al*, 1988).

In psychiatric research, it is essential to have uniform diagnostic criteria and to use instruments that are valid, sensitive, and reliable (Hamilton, 1982). The introduction of semi-structured clinical interviews (Spitzer *et al*, 1964; Wing *et al*, 1967) has greatly improved the rating of individual psychiatric symptoms and the psychiatric diagnosis derived from them. In addition, this method has proven validity, enabling satisfactory comparison of data and replication of studies by trained interviewers. Another advantage is the derivation of computer diagnosis for use as a reliable standard (Spitzer & Endicott, 1968; Wing, 1974). Although extensively used in the derivation of psychiatric diagnosis, the semi-structured method has never been used in instruments to compare the effects of treatment or to measure changes with treatment.

There is a need to develop a brief, semi-structured, depression rating scale which is reliable and valid, and which is sensitive to the specific symptomatological changes of depression in the elderly. The purpose of this study was to derive and evaluate such an instrument.

Method

A total of 80 elderly in-patients, 50 women (mean age 73 years; range 65–88) and 30 men (mean age 72; range 61–84) were studied. They all had a DSM–III diagnosis of major affective disorder, depression (American Psychiatric Association, 1980), as well as fulfilling the Research Diagnostic Criteria for depression (Spitzer *et al*, 1975). Other psychiatric diagnoses were excluded by the Geriatric Mental State Schedule (GMSS) and the Hodgkinson Mental State Schedule.

Instruments

The GMSS is a semi-structured clinical interview of demonstrated reliability and validity which is used extensively, as in the US/UK diagnostics project (Copeland *et al*, 1976). Sixteen subscales (102 items) of the GMSS related to depressive symptoms comprise the item pool from which this depression scale has been developed. The 16 subscales are the following: worry, general anxiety, depression, hypochondriasis, tension, somatic dysfunction, phobias, autonomic symptoms, thinking difficulties, slowing, loneliness, guilt, irritability, interest, concentration, and insight.

The HRSD, the Beck Depression Inventory (BDI; Beck *et al*, 1961), and the 102 GMSS items were completed before and after 6 weeks of treatment with tricyclic antidepressants by trained clinicians. In addition, the Newcastle Diagnostic Index (Carney *et al*, 1965) was completed at admission, and a seven-point (1 = very much improved, 2 = much improved, 3 = slightly improved, 4 = no change, 5 = slightly worse, 6 = much worse, and 7 = very much worse) global scale of improvement was completed by both the patient and an independent clinician after treatment.

Procedure

Subjects were considered to be improved if they met three criteria: (a) a before-to-after decrease in their BDI score

https://doi.org/10.1192/bjp.164.4.522 Published online by Cambridge University Press

of at least 50%, (b) a before-to-after decrease in their HDS score of at least 50% or a score lower than 10 after the test, and (c) a clinician's rating of very much or much improved.

In all, 60 subjects were classified as improved and 20 were classified as not improved, and the two groups were compared on items from the GMSS. GMSS items that were the most discriminating between the two groups were retained for the GMS–Depression Scale (GMS–DS). Reliability and validity coefficients were then calculated for the GMS–DS.

Item analysis of GMSS

For each of the GMSS items, before-to-after differences were calculated; a score of greater than 0 was categorised as improvement, and a score of 0 was categorised as non-improvement. χ^2 and odds ratios (Hennekens & Buring, 1987) were then calculated, contrasting the improved and not improved groups on each GMSS item. Items were retained for the short-form depression scale if the resulting χ^2 was significant at the (0.001) level, and the 95% confidence intervals for the odds-ratio did not overlap 1.0. An odds ratio of 1 indicated no association between the item and improvement, whereas an odds ratio of greater than 1 indicated a positive association.

In total, 33 items met this criterion: 2/8 worry items, 1/1 general anxiety item, 3/19 depression items, 1/6 hypochondriasis items, 2/4 tension items, 7/8 somatic items, 2/3 thinking difficulties items, 5/8 slowing items, 2/4 loneliness items, 1/2 guilt items, 1/8 irritability items, 3/4 interest items, 2/2 concentration items, and 1/5 insight items. Table 1 presents the χ^2 significance values, odds ratios, and 95% confidence intervals for the odds ratios for each of these items.

These 33 items comprise the GMS–DS. The GMS–DS takes about 15 min to administer with a depressed, elderly population and has a range of 0–71.

Table 1
Statistical results for the 33 GMS–DS items

GMS–DS item	χ^2 significance	Odds ratio	95% confidence intervals of odds ratio	
1	*	15.54	1.95	123.71
2	*	5.44	1.80	16.42
3	**	8.25	2.57	26.38
4	**	59.00	6.78	512.69
5	**	23.22	5.40	99.73
6	**	32.66	8.10	131.69
7	**	7.93	2.09	30.00
8	**	17.11	4.46	65.57
9	**	59.00	6.78	512.69
10	**	12.00	3.63	39.58
11	**	6.41	2.10	19.55
12	**	17.11	4.46	65.57
13	*	14.52	1.82	115.68
14	*	7.57	2.32	24.61
15	**	35.44	6.72	186.82
16	**	7.66	2.48	23.68
17	**	39.33	4.49	344.33
18	**	29.00	5.51	152.50
19	**	29.00	5.51	152.50
20	**	35.44	6.72	186.82
21	**	9.00	2.52	32.06
22	**	10.52	3.29	33.58
23	**	15.54	3.61	66.76
24	*	24.84	3.12	197.82
25	*	9.00	1.91	42.23
26	*	5.44	1.80	16.42
27	**	9.75	3.04	31.21
28	**	19.00	4.43	81.38
29	**	17.00	4.94	58.48
30	**	45.00	8.98	225.33
31	*	7.53	1.91	29.62
32	**	19.00	4.43	81.38
33	**	35.44	6.72	186.82

*$P<0.001$; **$P<0.0001$.

Results

Reliability

Cronbach's alpha coefficient (Cronbach, 1951) for the 33 items of the GMS–DS was calculated with the post-test scores. An alpha coefficient of 0.95 indicated that the GMS–DS has strong internal consistency.

A Spearman–Brown split-half correlation coefficient was also calculated, and the resulting correlation coefficient of 0.92 further indicated that the GMS–DS has good internal reliability.

Test–retest correlations were not calculated because it was not expected that the depression scores would be stable over time in light of the treatment interventions before and after testing.

Convergent validity–level of depression

Table 2 presents Pearson product-moment correlations after the test and between the GMS–DS, and the BDI, HRSD, and the clinician's rating of level of depression. The GMS–DS was highly correlated with the HRSD, BDI, and the clinician's rating of severity of depression, indicating good convergent validity for the GMS–DS as a measure of severity of depression.

Convergent validity – improvement in level of depression

Table '3 presents Pearson product-moment correlations between before-and-after change scores for the GMS–DS, HRSD, BDI, and clinician's and patient's ratings of improvement. The GMS–DS was highly correlated with the BDI, HRSD, and clinician's and patient's rating of

Table 2
Pearson correlations of post-test depression scores

	HDS	BDI	Clinician's rating of severity
GMS–DS	0.91	0.86	0.84
HRSD	–	0.86	0.86
BDI	–	–	0.74

All $P<0.001$.

https://doi.org/10.1192/bjp.164.4.522 Published online by Cambridge University Press

Table 3
Pearson correlations of before-to-after difference scores

	HDS	BDI	Clinician's rating of improvement	Patient's rating of improvement
GMS-DS	0.79	0.76	0.89	0.85
HRSD	–	0.81	0.81	0.73
BDI	–	–	0.73	0.66
Clinician's rating of improvement	–	–	–	0.91

All $P<0.001$.

improvement, indicating good convergent validity as a measure of change in level of depression.

Discriminant validity – level of depression

The GMS-DS was used to determine differences in categorised level of severity of depression of current episode as rated by trained clinicians before testing. The categories were mild, moderate, and severe, although none of the subjects in this study were rated as having mild severity of depression. Subjects rated as having a severe level of depression scored significantly higher ($t=4.27$, d.f. $=78$, $P<0.001$) on the GMS-DS (mean 48.8, s.d. 3.0; $n=30$) than did those rated as having a moderate level of depression (mean 44.4, s.d. 5.0; $n=50$).

Discriminant validity – sensitivity and specificity in predicting improvement

Sensitivity, specificity, and overall efficiency were calculated by the formulae outlined by Insel & Goodwin (1983). Figures 1 and 2 present the sensitivity, specificity, and efficiency of the GMS-DS scores at all possible cut-off points for after-test scores and before-to-after difference scores, respectively. Sensitivity represents the true-positive rate (percentage of improved group rated as improved), specificity the true-negative rate (percentage of not-improved group classified as not improved), and efficiency the overall percentage correctly classified.

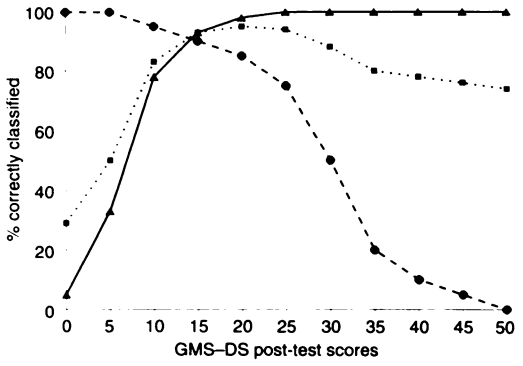

Fig. 1 Sensitivity (▲ true positives), specificity (---● true negatives), and efficiency (···■··· overall accuracy) of GMS-DS post-test scores.

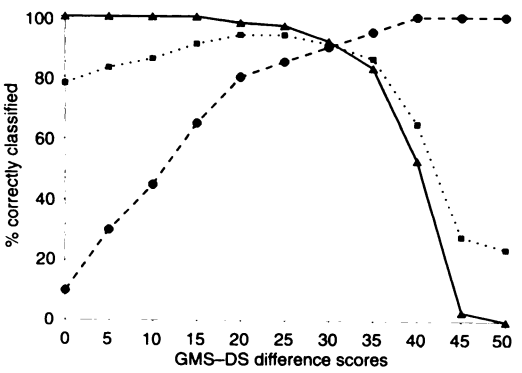

Fig. 2 Sensitivity (▲ true positives), specificity (---● true negatives), and efficiency (···■··· overall accuracy) of GMS depression scale before-to-after difference scores.

With after-test GMS-DS scores (Fig. 1), a cut-off of 18 resulted in a sensitivity of 97%, a specificity of 90%, and an efficiency of 95%. A range of 12–20 produced similar results. At 25, the true-positive rate was 100% although the true-negative rate dropped to 75%. A cut-off of 9 resulted in 100% of the not-improved group being classified as true-negative, while the true-positive rate dropped to 70%.

For before-to-after difference scores (Fig. 2), a cut-off of 30 resulted in a sensitivity of 92%, a specificity of 90%, and an efficiency of 91%, a range of 25–35 producing similar results. A cut-off of 18 resulted in a sensitivity, specificity, and efficiency of 100%, 80%, and 95%, respectively. A cut-off of 36 resulted in respective sensitivity, specificity, and efficiency of 78%, 100%, and 84%.

Figure 3 presents the GMS-DS before and after scores for the improved and not-improved groups. As evident from Fig. 1, both groups scored very high in level of depression before testing; in fact, close to the maximum possible score. Because of the extreme scores for this severely depressed population, some positive change over

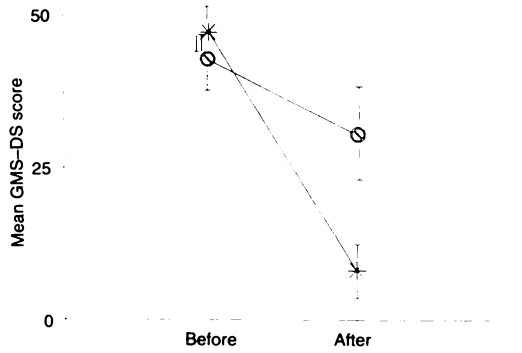

Fig. 3 Before and after depression scores on GMS-DS (✳ improved, $n=60$; ⊘ not improved, $n=20$; s.d.).

https://doi.org/10.1192/bjp.164.4.522 Published online by Cambridge University Press

time is to be expected, artefactually, as a function of regression to the mean. Analysis of the GMS–DS scores reveals that, while both groups improved, the improved group was significantly less depressed than the not-improved group ($t = 14.5$, d.f. = 78, $P < 0.001$) after testing.

Discussion

The results indicate that the semi-structured depression scale (GMS–DS) derived from the GMSS is a reliable and valid instrument for measuring severity of and change in depressive symptoms with treatment in the elderly population. In addition to correlation with standardised instruments such as the HRSD and the BDI, it shows even better correlation with the independent clinician's judgement of improvement as well as with that of the patient.

The GMS–DS comprises 33 items, and while this is more than most depression scales used to measure change, it is still very sparing of the interviewer's time. We felt that the inclusion of a sufficient number of items would greatly enhance the reliability of the instrument because the frequency of different depressive symptoms shows greater variability in the elderly than in the younger population. In addition, the increased number of items compensates for reduced range in the instrument (0–2 in most items), thus maintaining sensitivity without reduction of reliability.

Although most of the items included in the scale are core symptoms of depression seen in the younger population, certain symptoms such as irritability and indecisiveness appear to be more significant in this age group as changed symptoms. On the other hand, some symptoms noted to be very sensitive to change with treatment in younger patients were not included in the scale because the symptoms, although frequent, often continue with varying intensity as residual symptoms in the elderly in spite of clinical improvement. Such symptoms include guilt, pessimism, peripheral anxiety symptoms, dissatisfaction, and impaired self-image. This is consistent with previous evidence that these symptoms may be present in varying degrees in normal elderly subjects (Zung & Green, 1972). Even among the different manifestations of the same category of symptom, some symptoms appeared to be more sensitive to change than others. For example, the central manifestations of anxiety improved, while the peripheral anxiety symptoms such as trembling remained as residual symptoms. Excessive worry improved with treatment, but general pessimism remained. Irritability, autonomic symptoms, and phobic symptoms are said to be common in elderly depressed patients. Although the results confirmed this impression, these symptoms were not very sensitive to change with treatment. While specific hypochondriac symptoms such as headache and other pain remained residual in many, subjective distress was alleviated. Similarly, while loneliness remained as a symptom, it was less distressing. Psychomotor retardation has been eliminated as an item sensitive to change in some recent rating scales (Montgomery & Asberg, 1979) because it is thought to occur infrequently in younger depressed patients. This symptom occurred frequently in our sample and appeared to be very sensitive to change.

In conclusion, the GMS–DS has shown reliability and validity as a measure of both severity of depression and change in depressive symptoms for the elderly. The elimination of residual symptoms specific to the elderly depressed suggests that this is a useful instrument in clinical trials of treatment forms for an elderly, depressed population.

One important consideration is to what extent the demonstrated reliability and validity of the GMS–DS are sample specific. Validity coefficients often decrease in replication studies. There is a need for cross-validation of the GMS–DS with other, depressed, elderly populations to increase confidence in the psychometric properties of the instrument.

Appendix: Geriatric Mental State Schedule – Depression Scale (GMSS–DS): depression items sensitive to change with treatment

Worry

1. How much do you worry?

Worries a lot (i.e. about one or two things) 0 1 2 8 9
Is a worrier or worries about almost everything 0 1 2 8 9

2. Does this worrying bother you a lot? Is it unpleasant? (Can you stop yourself worrying?) Do the thoughts keep coming back?

Unpleasant worrying which keeps coming back or cannot be stopped 0 1 2 8 9

General anxiety

3. Do you get frightened? (Very anxious?) (Has that happened lately?) (What made you feel that way?)

Subjective fear or anxiety, out of proportion to the event, if any, that provoked the feeling 0 1 2 8 9

Depression

4. Have you been sad (depressed, miserable, in low spirits, blue) recently?

Depressed mood 0 1 2 8 9

https://doi.org/10.1192/bjp.164.4.522 Published online by Cambridge University Press

5. Have you felt like crying (wanted to cry) without actually weeping? (How often?)

Has felt like crying 0 1 2 8 9

6. Is the depression/crying/feeling like crying there most of the time? How long does it last? (Just a few hours at a time or longer than that?) How long have you had it? (Just a few hours at a time or longer than that?)

Depression, crying or feeling like crying lasts longer than just the occasional few hours 0 1 2 8 9

Hypochondriasis

7. How is your physical health? Is there anything about your body which bothers or upsets you? Are you in pain? Or is there any part of your body not working properly? (Would you say you are physically fit?)

Has a physical problem which causes emotional distress or worry 0 1 2 8 9

Tension

8. Do you get worn out (exhausted?) If no: what about towards the evening?

Gets worn out or exhausted during daytime or evening 0 1 2 8 9

9. Do you have difficulty in relaxing (resting)?

Difficulty in relaxing 0 1 2 8 9

Somatic dysfunction

10. What has your appetite been like? Do you enjoy your food? Have you been eating more or less than usual?

Diminution in the *desire* for food 0 1 2 8 9
Increase in the *desire* for food 0 1 2 8 9

11. Why is that? Has it been like that most days in the last month?

Poor appetite in the absence of known medical condition and without nausea 0 1 2 8 9

12. Have you lost any weight during the past three months?

Lost 10 lb (4.5 kg) or more over the past 3 months 0 1 2 8 9

13. Have you had trouble sleeping recently? (Have you taken anything to help you sleep?) How long has it been going on for? What used to happen?

Trouble with sleep or recent change in pattern 0 1 2 8 9

14. Have you had any difficulty falling asleep (getting off to sleep)? Do you lie awake for long periods of time (waiting for sleep)?

Difficulty in falling asleep. If tablets taken, rate what interviewee *feels* would have happened without them 0 1 2 8 9

15. Is your sleep interrupted during the night?

Sleep interrupted during the night. (Include waking up after an initial cat nap and not being able to sleep again for some lengthy time.) If tablets taken rate what interviewee *feels* would happen without them 0 1 2 8 9

16. Have you recently been waking up early in the morning and found it impossible to get back to sleep? What time would that be? How often has it happened?

Awakens about two hours or more before normal time of awakening and *cannot* get back to sleep, most nights for at least two weeks in the last month 0 1 2 8 9

Thinking difficulties

17. Do your thoughts get mixed up (muddled?) (so that you cannot get them sorted out?) (Can you think clearly (straight)?) (How long has that bothered you? How often?)

Feeling of being muddled 0 1 2 8 9

18. Do you find it difficult to make up your mind (to make decisions)? (How long has that bothered you? How often?)

Feels indecisive 0 1 2 8 9

Slowing

19. Have you had too little energy (to do things you want to do)? How long have you had that for? Are you like that most days?

Listlessness or subjective restriction of energy 0 1 2 8 9

20. Have you been doing more, less, or about the same as usual?

Doing less than usual 0 1 2 8 9

21. Did this (slowing, loss of energy, reduced activity) start in the last three months or perhaps get worse in the last three months?

Started or became worse in the last few months 0 1 2 8 9

22. What about when someone visits you or you have to go out? Does that make any difference?

Does not lift with usually pleasant activities 0 1 2 8 9

23. Have you actually been sitting around a lot (or spending more time in bed than usual) because of lack of energy?

Sits or lies around because of lack of energy 0 1 2 8 9

Loneliness

24. Do you feel lonely?

Admits to feeling lonely 0 1 2 8 9

25. Does it bother you very much (make you feel depressed?) Can you get out of it?

Feels lonely and cannot turn away from it 0 1 2 8 9
Bothered or depressed by current loneliness 0 1 2 8 9

https://doi.org/10.1192/bjp.164.4.522 Published online by Cambridge University Press

Guilt

26. Do you tend to blame yourself for anything or feel guilty about anything? What? (Do you mean you actually feel worthless?) (How long have you felt like this?) Is it reasonable?

Obvious excessive guilt or self blame over past and present peccadilloes. (Do not include justifiable or minor self-blame.) 0 1 2 8 9

Mentions regrets about past which may or many not be justifiable 0 1 2 8 9

Irritability

27. Do you get angry with yourself?

Gets angry with self 0 1 2 8 9

Interest

28. How is your interest in things? (Do you keep up your interests?)

Has less interest in things than is usual for him/her 0 1 2 8 9

29. When did you notice this loss of interest/enjoyment? When did it start? Has it been present recently? For how long? Is it there most days?

Falling off of interest/enjoyment has occurred over the last 3 months 0 1 2 8 9

30. Is it that you are too depressed or nervous?

Too depressed or nervous 0 1 2 8 9

Concentration

31. Can you concentrate on a television (radio, film) programme? (Can you watch it (listen to it) all the way through?)

Difficulty in concentrating on entertainment 0 1 2 8 9

32. Do you read? Can you concentrate on something you read? (Can you read it right through?)

Difficulty in concentrating on reading 0 1 2 8 9

Insight

33. In general, how happy would you say you are: very happy, fairly happy, not very happy, or not happy at all?

0 = very happy, 1 = fairly happy,
2 = not very happy, 3 = not happy at all 0 1 2 3 8 9

Coding for item ratings is as follows (*unless otherwise specified*):

0 = No (or 'normal')
1 = Yes (or 'abnormal') but mild to moderate intensity, infrequent or fleeting
2 = Yes (or 'abnormal') and severe, frequent or persistent
8 = No reply elicited OR question not understood OR reply inaudible, inappropriate, or incoherent OR rating uncertain
9 = question not asked or inapplicable.

References

AMERICAN PSYCHIATRIC ASSOCIATION (1980) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn) (DSM-III). Washington, DC: APA.

BECK, A. T., WARD, C. H., MENDLESON, M., *et al* (1961) An inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–571.

CARNEY, M. W. P., ROTH, M. & GARSIDE, R. F. (1965) The diagnosis of depressive syndromes and the prediction of ECT response. *British Journal of Psychiatry*, **111**, 659–674.

COPELAND, J. R. M., KELLEHER, M. J., KELLETT, J. M., *et al* (1976) A semi-structured clinical interview for the assessing of diagnosis and mental state in the elderly; the Geriatric Mental State Schedule. *Psychological Medicine*, **6**, 439–449.

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

HAMILTON, M. (1967) Development of a rating scale for primary depressive illness. *British Journal of Psychiatry*, **6**, 278–296.

——— (1982) Symptoms and assessment of depression. In *Handbook of Affective Disorders*. (ed. E. S. Paykel), pp. 1–11. New York: Guilford Press.

HENNEKENS, C. & BURING, J. (1987) *Epidemiology in Medicine*. Boston: Little, Brown.

INSEL, T. & GOODWIN, F. (1983) The dexamethasone suppression test: promises and problems of diagnostic laboratory tests in psychiatry. *Hospital and Community Psychiatry*, **34**, 1131–1138.

KOENIG, H. G., MEADOR, K. G., COHEN, H. J., *et al* (1988) Self-rated depression scale and screening for major depression in the older hospitalized patient with medical illness. *Journal of the American Geriatrics Society*, **36**, 699–704.

MONTGOMERY, S. A. & ASBERG, M. (1979) A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, **134**, 382–389.

POST, F. (1982) Affective disorders in old age. In *Handbook of Affective Disorders* (ed. E. S. Paykel), pp. 392–402. New York: Guilford Press.

SPITZER, R. L., FLEISS, J. L., BURDOCK, E. I. & HARDESTY, A. S. (1964) The mental status schedule; rationale, reliability and validity. *Comprehensive Psychiatry*, **5**, 384–395.

——— & ENDICOTT, J. (1968) Diagno: a computer program for psychiatric diagnosis utilizing the differential diagnostic procedure. *Archives of General Psychiatry*, **18**, 746–756.

———, ——— & ROBINS, E. (1975) *Research Diagnostic Criteria, Biometrics Research*. New York State Department of Mental Hygiene.

WING, J. K. (1974) *Measurement and Classification of Psychiatric Symptoms*. Oxford: Oxford University Press.

———, BIRLEY, J. L. T., COOPER, J. E., *et al* (1967) Reliability of a procedure for measuring and classifying present psychiatric state. *British Journal of Psychiatry*, **113**, 499–515.

ZUNG, W. K. & GREEN, R. L. (1972) Detection of affective disorders in the aged. In *Psychopharmacology and Aging* (eds C. Eisdofer & W. E. Fann), pp. 213–224. New York: Plenum Press.

* A. V. Ravindran, MB, MRCPsych, MSc, FRCP(C), *Associate Professor, Department of Psychiatry, University of Ottawa, Royal Ottawa Hospital, 1145 Carling Avenue, Ottawa, Ontario, Canada K1Z 7K4*; K. Welburn, PhD, *Clinical Psychologist, Royal Ottawa Hospital*; J. R. M. Copeland, MD, FRCP, FRCPsych, *Professor and Chairman, Department of Psychiatry, University of Liverpool, Royal Liverpool Hospital, P.O. Box 147, Liverpool L69 3BX*

*Correspondence

(*First received June 1992, final revision May 1993, accepted July 1993*)

https://doi.org/10.1192/bjp.164.4.522 Published online by Cambridge University Press