Psychological Medicine, 1997, **27**, 967–971.

BRIEF COMMUNICATION

Performance on the delayed word recall test (DWR) fails to differentiate clearly between depression and Alzheimer's disease in the elderly

R. E. O'CARROLL,[1] S. CONWAY, A. RYMAN AND N. PRENTICE

From the Department of Psychology, University of Stirling; and MRC Brain Metabolism Unit, Royal Edinburgh Hospital, Morningside Park, Edinburgh

ABSTRACT

Background. The differential diagnosis of early dementia of the Alzheimer's type from depression in the elderly is often made difficult by the presence of significant memory impairment in depressed patients. The Delayed Word Recall test (DWR) was developed to facilitate the early diagnosis of Alzheimer's disease. The DWR involves: (*a*) repeated elaborate encoding of ten separate words; (*b*) a filled delay; (*c*) delayed free recall. A recognition memory test has also been recently developed. The available evidence suggests impressive sensitivity and specificity when the DWR has been used to separate patients with early Alzheimer's disease from very well matched controls.

Methods. In the present study, the DWR was evaluated with regard to its ability to separate a group of 50 patients with early Alzheimer's disease from 50 elderly patients with major depression in a between-subjects experimental design.

Results. For both free recall and recognition indices, the between-group overlap was large. Using recommended cut-off scores for the detection of Alzheimer's disease, 44% of the depressed patients would have been misclassified as demented based on their free recall scores, and 48% of the depressed patients would have been misclassified on the basis of their recognition scores.

Conclusion. We conclude that the DWR is not specific enough to clearly distinguish patients with early Alzheimer's disease from elderly patients with major depression.

INTRODUCTION

The memory impairments associated with Alzheimer's disease are well documented (Albert *et al.* 1989). In two recent studies by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), the investigators reported that delayed verbal recall was a highly sensitive indicator of early Alzheimer's disease (Morris *et al.* 1989; Welsh *et al.* 1992). In Alzheimer's disease there is a rapid rate of forgetting within the first 5–10 min following acquisition (Butters *et al.* 1988; Hart *et al.* 1988), but a relatively normal rate of forgetting thereafter (Corkin *et al.* 1984; Kopelman, 1985). Knopman & Ryberg (1989) highlighted the fact that elaborative encoding provides a substantial benefit to normal elderly individuals, but not to patients with Alzheimer's disease. They concluded that the optimal memory test procedure for the early diagnosis of dementia due to Alzheimer's disease should involve: (*i*) a study phase in which subjects were required to engage in elaborative encoding; (*ii*) a delay interval; and (*iii*) a test phase which involved free recall. Consequently, they developed the Delayed Word Recall (DWR) test.

[1] Address for correspondence: Dr R. E. O'Carroll, Department of Psychology, University of Stirling, Stirling FK9 4LA.

This test, briefly, comprises 10 nouns presented on cards one at a time (chimney, salt, harp, button, meadow, train, flower, finger, rug and book). For each word, the subject is asked to read out the word, to try and remember it, and then to make up a sentence using the word. When all 10 words have been presented, the complete procedure is repeated, with subjects instructed to produce an alternative sentence. Following a 5 min filled delay, free recall for the 10 words list is then tested. Based on the delayed free recall cut-off score of less than or equal to two out of 10 implying Alzheimer's disease, Knopman & Ryberg (1989) reported an overall predictive accuracy of 95% in 28 Alzheimer patients and 55 elderly control subjects. Only one healthy subject scored less than two on the DWR test, and only three early Alzheimer patients scored more than two out of 10. In a follow-up study, Coen *et al*. (1996) compared 42 mild to moderately demented Alzheimer patients with 42 age-matched healthy controls and found that the DWR recall measure achieved 98% sensitivity, specificity and overall accuracy. In an extension to the original DWR, Coen *et al*. (1996) included a four choice recognition memory test and using a cut-off score of less than or equal to nine out of 10 implying Alzheimer's disease, they reported that the recognition measure yielded a sensitivity of 98%, a specificity of 95% and overall accuracy of 96%. Taken together, the results of the Knopman & Ryberg (1989) and Coen *et al*. (1996) studies suggest that the DWR is an extremely promising measure to aid in the early detection of Alzheimer's disease.

However, in both of these studies, the specificity and sensitivity of the DWR was determined in distinguishing a well-defined Alzheimer group from a well matched healthy control sample. In clinical practice, the most difficult differential diagnoses are of early Alzheimer's disease from depressive disorders. This has been described by Lezak (1995) as 'probably the knottiest problem of differential diagnosis' (p. 327). In a previous study, it was shown that while the discrepancy between pre-morbid intellectual measures and mean Wechsler Memory Scale – Revised indices clearly separated Alzheimer patients from healthy controls, the overlap between Alzheimer and elderly depressed patients was large (O'Carroll *et al*. 1994). The aim of the current investigation was to determine the ability of the DWR test (both free recall and recognition) to differentiate between patients with early Alzheimer's dementia and elderly patients suffering from a major depressive episode.

METHOD

Fifty patients with dementia of the Alzheimer-type were recruited. All met DSM-IV criteria for dementia and NINCDS-ADRDA criteria for probable Alzheimer's disease (McKhann *et al*. 1984). The criteria include steadily progressing dementia with dysmnesia as the initial feature with no history suggestive of other types of dementia, absence of focal neurological signs, no evidence of hypertension or other cardiovascular abnormalities, and normal haematological and biochemical investigations. The DAT subjects had a Mean Mini-Mental State Examination (Folstein *et al*. 1975) score of 16·1 (5·2).

Fifty patients who fulfilled DSM-IV criteria for a major depressive episode with or without melancholic or mood congruent psychotic features were also entered into the study. In addition, the GMS-AGECAT semi-structured diagnostic interview schedule was used to confirm diagnosis (Copeland *et al*. 1986). Mean severity of depression as assessed by the 21-item Hamilton Depression Rating Scale (Hamilton, 1960) was 23·2 (5·3) and for the 17-item scale, 20·8 (4·5).

All subjects were assessed using the National Adult Reading Test – Revised (Nelson & Willison, 1991) to provide an estimate of pre-morbid intellectual level, and years of full-time education was recorded. Patients were administered the Delayed Word Recall test following Knopman & Ryberg (1989). In addition, the multiple choice delayed recognition memory test was administered, as described by Coen *et al*. (1996).

RESULTS

Demographic details of the two subject groups were as follows: depressed patients – age 73·9 (6·4) years, DAT patients 77·0 (8·0) years, $t = 2{\cdot}13$, $P = 0{\cdot}04$; education – depressed patients 10·2 (2·6) years, DAT patients 10·3 (2·3) years, $t = 0{\cdot}25$, $P = 0{\cdot}80$; NART IQ – depressed patients 96·9 (14·5) years, DAT patients 97·8 (15·2) years, $t = 0{\cdot}27$, $P = 0{\cdot}79$; MMSE total –

https://doi.org/10.1017/S0033291796004679 Published online by Cambridge University Press

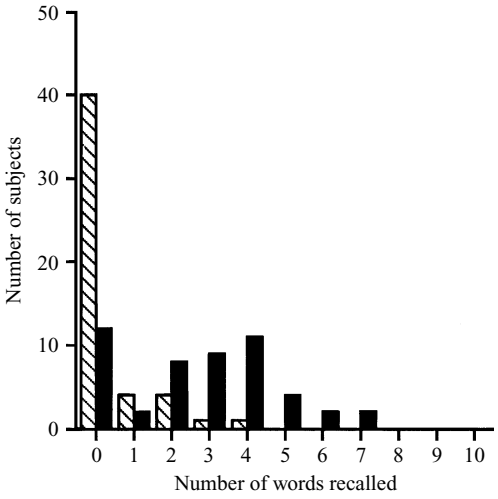

FIG. 1. DWR recall results: using < 3 recommend cut-off only 2 (4%) DAT (▨) subjects would be misclassified, but 22 (44%) DEP (■) subjects would be classed as DAT.

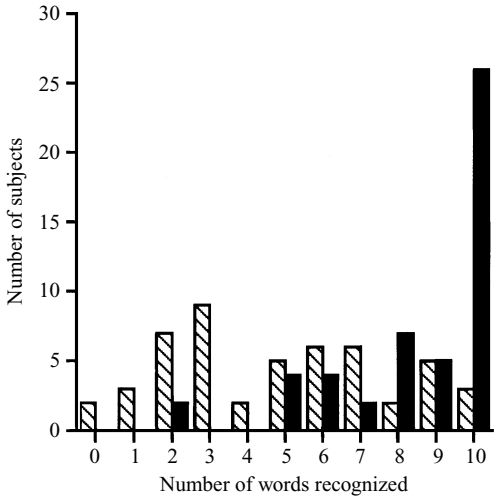

FIG. 2. DWR recognition results: using < 10 recommended cut-off only 3 (6%) DAT (▨) subjects would be misclassified, but 24 (48%) DEP (■) subjects would be classed as DAT.

depressed patients 25·8 (3·8) years, DAT patients 16·1 (5·2) years, $t = 10·6$, $P = 0·00$; sex – depressed patients 19 males, 31 females, DAT patients 13 males, 37 females, $\chi^2 = 1·66$, $P = 0·19$. As there was a slight but significant age difference between the groups, age was entered as a co-variate in the comparison of DWR indices. DWR free recall depressed mean = 2·70 (2·00), DAT mean = 0·38 (0·88), $F = 49·2$, $P = 0·00$; DWR recognition depressed mean = 8·46

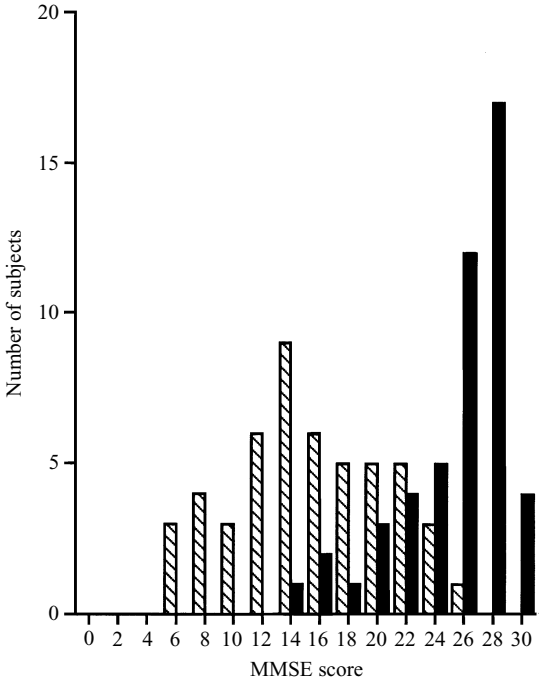

FIG. 3. MMSE results: using < 24 recommended cut-off 4 (8%) DAT (▨) subjects would be misclassified, but 12 (22%) DEP (■) subjects would be classed as DAT.

(2·14), DAT mean = 4·92 (2·86), $F = 45·9$, $P = 0·00$. The correlations between DWR recall and recognition and other variables for the total sample of 100 subjects were as follows: age *versus* recall, $r = -0·29$ ($P < 0·01$); age *versus* recognition, $r = 0·13$ (NS); 17-item Hamilton *versus* recall, $r = 0·02$ (NS); 17-item Hamilton *versus* recognition, $r = 0·10$ (NS); MMSE total *versus* recall, $r = 0·65$ ($P < 0·01$); MMSE total *versus* recognition, $r = 0·73$ ($P < 0·01$); NART IQ *versus* recall, $r = 0·08$ (NS); NART IQ *versus* recognition, $r = 0·26$ ($P = 0·01$); DWR recognition *versus* DWR recall, $r = 0·57$ ($P < 0·01$).

The breakdown of test scores by patient group are presented in Figs. 1–3. For DWR recall, sensitivity = 96%, specificity = 69% and overall accuracy = 76%. For DWR recognition, sensitivity = 94%, specificity = 66% and overall accuracy = 73%. For the MMSE, sensitivity = 92%, specificity = 79% and overall accuracy = 84%. The results of a separate discriminant function analysis for the three measures are as follows: DWR recall eigen = 0·57, $\chi^2 = 44·3$, overall correct classification = 80%; DWR

https://doi.org/10.1017/S0033291796004679 Published online by Cambridge University Press

recognition eigen = 0·50, $\chi^2 = 39{\cdot}6$, overall correct classification = 74%; MMSE eigen = 1·15, $\chi^2 = 73{\cdot}9$, overall correct classification = 82%.

DISCUSSION

There were highly significant differences in mean DWR recall and recognition scores when the depressed and Alzheimer groups were compared. However, inspection of Figs. 1 and 2 reveal that the between-group overlap was large i.e. the separation was poor. This result is in stark contrast to the excellent separation observed on DWR indices in the comparison of Alzheimer patients and matched healthy controls (Knopman & Ryberg, 1989; Coen *et al*. 1996). In fact, the best between-group separation in the present study was found by using the MMSE, where an overall accuracy of 84% was achieved. Although the DWR is sensitive in detecting the majority of DAT cases, its specificity is poor. It is becoming increasingly clear that major depression in the elderly is associated with significant and marked impairment of memory functioning (Abas *et al*. 1990; Austin *et al*. 1992; Robbins *et al*. 1992; Moffoot *et al*. 1994; O'Carroll *et al*. 1994; Ilsley *et al*. 1995) and that simple reliance on memory as a primary diagnostic criterion can lead to mis-diagnosis of a significant number of depressed patients as demented, with the consequence of depressive illnesses left untreated.

The better separation provided by the MMSE may be a consequence of it tapping non-mnemonic cognitive functions which are less impaired in depression e.g. praxis and language. However, the overlap on MMSE scores also makes it clearly inadequate for diagnostic purposes in the individual clinical case. Furthermore, as in all studies of this type, the comparison groups were very clearly defined, and it is in cases of diagnostic doubt that neuropsychological measures, such as the DWR test, are required to demonstrate their clinical usefulness.

It could be argued that the memory impairment demonstrated by the depressed group was not a consequence of depression, but was due to antidepressant medications that the majority of the depressed patients were taking. However, the evidence, such as it is, suggests that after short-lived acute effects, which are attributable to sedation in depressed patients, antidepressant medication tends to improve, not impair, cognitive function (see review by Thompson, 1991). In any respect, if a neuropsychological measure such as the DWR is to prove useful in clinical practice, it is extremely unlikely that it would be used on significant numbers of drug-naive or drug-free patients.

In summary, we have carried out a large scale comparison of DWR recall and recognition measures in early Alzheimer and elderly depressed patients. We conclude that the between-group overlap on DWR indices is of such a magnitude as to render the measure inadequate for the differential diagnosis of depression *versus* early Alzheimer's disease in elderly patients.

We would like to acknowledge the help of Guy Goodwin, Klaus Ebmeier, Catherine Murray and Margaret Van Beck in the present study. We would also like to thank Norma Brearley for her careful preparation of the manuscript. Finally, we would like to thank Robert Coen for kindly providing us with a copy of the DWR recognition test.

REFERENCES

Abas, M. A., Sahakian, B. J. & Levy, R. (1990). Neuropsychological deficits and CT scan changes in elderly depressives. *Psychological Medicine* **20**, 507–520.

Albert, M. S., Moss, M. B. & Milberg, W. (1989). Memory testing to improve the differential diagnosis of Alzheimer's disease. In *Alzheimer's Disease and Related Disorders* (ed. K. Iqbal, H. M. Wisniewski and B. Winblad), pp. 55–69. Alan R. Liss: New York.

Austin, M.-P., Ross, M., Murray, C., O'Carroll, R. E., Ebmeier, K. P. & Goodwin, G. M. (1992). Cognitive function in major depression. *Journal of Affective Disorders* **25**, 21–30.

Butters, B., Salmon, D. P., Cullum, C. M., Cairns, P., Troster, A. I. & Jacobs, D. (1988). Differentiation of amnesic and demented patients with the Wechsler Memory Scale – Revised. *Clinical Neuropsychologist* **2**, 133–148.

Coen, R. F., Swanwick, G. R., Maguire, C., Kirby, M., Lawlor, B., Walsh, J. B. & Coakley, D. (1996). Memory impairment in Alzheimer's disease: replication and extension of the delayed word recall (DWR) test. *Irish Journal of Psychological Medicine* **13**, 55–58.

Copeland, J. R. M., Dewey, M. E. & Griffiths-Jones, H. M. (1986). Psychiatric case nomenclature and a computerised diagnostic system for elderly subjects: GMS and AGECAT. *Psychological Medicine* **16**, 89–99.

Corkin, S., Growdon, J., Nissen, M. J., Huff, F. J., Freed, D. M. & Sagar, H. J. (1984). Recent advances in the neuropsychological study of Alzheimer's disease. In *Advances in Basic Research and Therapies* (ed. R. J. Wurtman, S. Corkin and J. H. Growdon), pp. 75–94. Centre for Brain Sciences and Metabolism Trust: New York.

Folstein, M. F., Folstein, S. E. & McHugh, P. R. (1975). Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician *Journal of Psychiatric Research* **12**, 189–198.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* **23**, 56–62.

Hart, S., Smith, C. M. & Swash, M. (1988). Word fluency in patients with early dementia of Alzheimer type. *British Journal of Clinical Psychology* **27**, 115–124.

Ilsley, J., Moffoot, A. & O'Carroll, R. (1995). An analysis of memory dysfunction in major depression. *Journal of Affective Disorders* **35**, 1–9.

Knopman, D. S. & Ryberg, S. (1989). A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Archives of Neurology* **46**, 141–145.

Kopelman, M. D. (1985). Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome. *Neuropsychologia* **23**, 623–638.

Lezak, M. (1995). *Neuropsychological Assessment*, *3rd edn*. Oxford University Press: Oxford.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology* **34**, 939–944.

Moffoot, A. P. R., O'Carroll, R. E., Bennie, J., Carroll, S., Dick, H., Ebmeier, K. P. & Goodwin, G. M. (1994). Diurnal variation of mood and neuropsychological function in major depression with melancholia. *Journal of Affective Disorders* **32**, 257–269.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D. & Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part 1. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**, 1159–1165.

Nelson, H. E. & Willison, J. R. (1991). *The Revised National Adult Reading Test – Test Manual*. NFER-Nelson: Windsor.

O'Carroll, R. E., Curran, S. M., Ross, M., Murray, C., Riddle, W., Moffoot, A. P. R., Ebmeier, K. P. & Goodwin, G. M. (1994). The differentiation of major depression from dementia of the Alzheimer-type using within-subject neuropsychological discrepancy analysis. *British Journal of Clinical Psychology* **33**, 23–32.

Robbins, T. W., Joyce, E. M. & Sahakian, B. J. (1992). Neuropsychology and imaging. In *Handbook of Affective Disorders*, *2nd edn*. (ed. E. S. Paykel), pp. 289–309. Churchill Livingstone: Edinburgh.

Thompson, P. J. (1991). Antidepressants and memory: a review. *Human Psychopharmacology* **6**, 79–90.

Welsh, K. A., Butters, N., Hughes, J., Mohs, R. & Heyman, A. (1992). Detection and staging of dementia in Alzheimer's disease. Use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer's Disease. *Archives of Neurology* **49**, 448–452.