Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale

HARM W J VAN MARWIJK

PAUL WALLACE

GEERTRUIDA H DE BOCK

JO HERMANS

ADRIAN A KAPTEIN

JAN D MULDER

SUMMARY

Background. *Many scales have been developed to assess depression, but they are often too lengthy to be of practical use in general practice consultations.*

Aim. *A study was undertaken to investigate the feasibility, reliability and diagnostic value of the geriatric depression scale and its shorter versions for screening in general practice.*

Method. *A total of 586 consecutive consulting patients aged 65 years and over were studied in nine general practices in the west of the Netherlands (13 doctors). The 30-item version of the geriatric depression scale was compared with the diagnostic interview schedule as a reference test.*

Results. *The reference test indicated a major depression in six patients while 27 patients had a dysthymic disorder (that is, a chronic mild depression). Five per cent of patients required help for 50% of the questions on the geriatric depression scale. The diagnostic value of the 30-item, 15-item, 10-item and four-item versions did not differ significantly, but the one-item version performed no better than chance. Two items discriminated best between patients who were and who were not depressed (P<0.05), only one of which was included in a previously proposed four-item version of the scale. The reliability of the proposed four-item version was 0.64, the reliability of the other versions ranging from 0.70 to 0.87.*

Conclusion. *The results for the different versions of the geriatric depression scale suggest the use of a 10-item or a four-item version. For practical purposes, the smallest subset would be the most desirable: the four-item version. These scales may be better suited for exclusion rather than inclusion purposes. The feasibility of screening for depression in elderly people in a general practice setting is discussed in the light of the results of the study.*

H W J van Marwijk, MD, general practitioner; G H de Bock, MA, psychologist; and J D Mulder, MD, PhD, professor, Department of General Practice, Leiden University. J Hermans, PhD, biostatistician, Department of Statistics, Leiden University. A A Kaptein, PhD, psychologist, Department of Psychiatry, Leiden University, Leiden, Netherlands. P Wallace, MSc, FRCGP, professor, Department of General Practice and Primary Care, Royal Free Hospital School of Medicine, London.

Submitted: 21 March 1994; accepted: 25 August 1994.

Keywords: depression; screening; diagnosis; assessment techniques.

Introduction

DEPRESSION is an important health problem. Depressive symptoms, with or without depressive disorder, have been associated with negative outcomes such as diminished functioning or well being, and increased mortality.[1,2] A number of scales have been developed to assess depression. Examples of these are the self-care(D),[3] and the geriatric depression scale.[4]

The geriatric depression scale is a self-report instrument with 30 yes/no questions (rated 1/0). It identifies clinical depression according to the *Diagnostic and statistical manual of mental disorders,* third edition.[5,6] The 30-item version of the geriatric depression scale is considered a useful but somewhat lengthy screening instrument for depression in elderly people. Its feasibility and diagnostic value have been demonstrated in a variety of settings.[7] The reliability and validity of the Dutch translation of the geriatric depression scale appear satisfactory.[8]

A shorter 15-item version of the geriatric depression scale was developed in the United States of America (Appendix 1).[9] It retains the diagnostic value of the longer scale, while considerably shortening the time required for its administration.[10] A general practice package of health checks for people aged 75 years and over containing this 15-item version is now available for the routine screening of this group in the United Kingdom.[11]

While advocated for general use (for example by the Royal College of Physicians[12]) the 30- and 15-item versions of the geriatric depression scale are still thought to be too lengthy and therefore cumbersome for use in general practice. A version with fewer than 15 items would be seen as desirable by many general practitioners. One research team has derived versions with 10, four, and one items.[9,13]

A study was undertaken to determine the characteristics of the 30-item geriatric depression scale in general practice. Analyses were carried out in order to answer the following questions: What is the feasibility of using different subsets of questions from the geriatric depression scale as screening instruments for depression in elderly general practice patients? What is their diagnostic value (that is, sensitivity, specificity and negative and positive predictive value)? Can a previously proposed selection of 10, four, and one questions from the geriatric depression scale be confirmed? Are the subsets reliable (that is, internally consistent)?

Method

The diagnostic interview schedule was used for obtaining reference diagnoses of major depression and dysthymia (that is, a chronic mild depression).[14] This schedule is a widely accepted, structured psychiatric interview that was designed for use by trained lay interviewers. For the study, 11 interviewers received training in the use of appropriate sections (somatizing disorders, affective disorders, schizophrenia and cognitive impairment). The manual of mental disorders guided which sections were to be chosen.[15] The diagnosis of cognitive impairment is not a dia-

gnosis according to the manual, but is based on the mini-mental state examination. The mini-mental state examination forms an integral part of the diagnostic interview schedule.

Fourteen general practitioners in Leidschendam were invited to take part. Leidschendam is located near the Hague, in the densely populated western part of the Netherlands.

Eighty consecutive patients aged 65 years and over consulting in the surgery or at home were invited by each doctor to take part in the study. From February to April 1992 the practice assistants in each practice made lists of all these patients. Patients were excluded if they were assessed by the general practitioner or the interviewer as not capable of participating or if they declined to take part. During the consultation the general practitioner mentioned the research project, handed the patient a letter describing the project, and asked him or her to fill in a Dutch version of the 30-item geriatric depression scale at home. All the participating patients were subsequently visited by an interviewer. If requested by a patient, assistance in completing the questionnaire was provided by the interviewer at the visit. The interviewers used the diagnostic interview schedule to determine the six-month prevalence rates of major depression and dysthymic disorder.

Statistics

Data from the diagnostic interview schedule were analysed using the official Dutch version of the computer diagnostic programme.[15] Several checks were made to ensure accuracy of data collection and subsequent data entry. Data were entered using *SPSSPC+* data entry software. Double entry was performed on a random sample of 5% of the data, revealing accurate entry.

To determine the feasibility of administration of the geriatric depression scale, the number of times each patient needed help with an individual item was assessed.

To compare the diagnostic value of the different versions against the diagnostic interview schedule, receiver operating characteristic curves were used.[16] These allow the examination of a test's ability to discriminate between two states, regardless of the cut-off point selected. The area under the curve was used as a quantitative measure of test performance.[17,18] The differences between areas under the curves were calculated using 95% confidence intervals; a perfect diagnostic test will have an area under the curve equal to one, while a test with no diagnostic value will have an area equal to 0.5, which represents chance.

To determine whether these data support the choice of items proposed by D'Ath and colleagues,[9] a forward stepwise logistic regression was performed. At each step in the analysis, the item which discriminated most between depressed and non-depressed patients according to the reference interviews was chosen and the remaining items reanalysed. In this way, the questions for the 10-, four- and one-item versions of the questionnaire were selected. Cronbach's alpha was used for internal consistency reliability, where a value of 1.0 indicates complete internal consistency and a value of zero indicates a lack of internal consistency. Statistical analyses were performed with *SPSSPC+*, version 4.0.

Results

Thirteen general practitioners from nine practices participated in the study. Eleven general practitioners were involved in undergraduate or postgraduate medical training, or in research in general practice. Five doctors worked in a single-handed practice, the others in two-partner practices. Most (11) had at least 10 years' experience in general practice. The mean practice list size for the 13 general practitioners was 2250 patients (range 1400–3000). The mean list size in the Netherlands is 2350 patients.

Of 706 patients requested to participate, 120 were excluded. The general practitioners considered 20 patients not able to participate because they were cognitively impaired (six patients), unable to read or write (three), they did not understand the procedure (10), or were too depressed (one). Twenty five patients were otherwise excluded: 16 patients could not be contacted by the interviewer, four were too ill, and two had language problems (no reason given for three patients). Fifty patients refused to participate. There were incomplete data in 25 cases (the geriatric depression scale was incomplete in 13 cases and the diagnostic interview schedule in 12 cases).

The doctors spontaneously mentioned depression in six of the 50 patients who refused to participate. The 120 patients excluded from the study (mean age 76.3 years, range 66–92 years) were a mean of 2.8 years older than the 586 participants (mean age 73.5 years, range 65–94 years; $P<0.001$). Proportionally more women (83) than men were among the non-participants than the participants (349) (69.2% versus 59.6%, $P<0.05$).

Data from 586 patients were available for analysis (237 men and 349 women; 351 patients were aged between 65 and 74 years, and 235 were aged between 75 and 94 years). The age–sex distribution was representative of the parent population. A total of 390 patients were married or living with a partner, and 196 were single, widowed or divorced. Most patients (495) lived in independent housing while 91 lived in homes for elderly people. There were 464 practice visits and 122 home visits.

The mini-mental state examination (part of the diagnostic interview schedule) scores indicated mild cognitive impairment in 18 patients, mild or possibly severe cognitive deficit in one patient, and no cognitive impairment in 567 patients.

Feasibility

A total of 185 patients (31.6%) needed help with one, two or three questions on the geriatric depression scale; 101 (17.2%) needed help with three or more questions. In all versions the percentage of patients requiring help was the same (between 5% and 6%). Patients who required help needed help for 50% of the questions in all versions.

Diagnostic value

According to the diagnostic interviews, 33 patients had a positive reference test: six patients had evidence of a major depression in the last six months, and 27 of a more chronic dysthymic disorder. The estimated prevalence of recent major depression and dysthymic disorder was therefore calculated to be 5.6% (33/586).

The frequency of scores on the 30-item version of the geriatric depression scale among depressed and normal patients is shown in Figure 1. The distribution of scores in the depressed group is clearly to the right of the normal group. Figure 2 shows receiver operating characteristic curves for the different versions of the geriatric depression scale. The curve for a perfect test would approach the upper left hand corner; even the 30-item version of the test is not perfect. The actual questions selected in the 10-, four- and the one-item versions are shown in Appendix 1.

The prevalences, sensitivity, specificity and positive and negative predictive values according to the reference standard, the diagnostic interview schedule, are shown in Table 1. For screening purposes, a high negative predictive value is most desirable. For example, a cut-off point of two on the 15-item version is associated with a sensitivity of 76%, a specificity of 53%, and a positive and negative predictive value of 9% and 97%, respectively. The use of the 15- or 10-item version where the cut-off point is three changes a prior probability of disease of 6% (that is, the prevalence of major depression and dysthymia) into a probability after the test (that is, a positive predictive value) of 13% or 15%, respectively. Few patients who are depressed according to the reference standard are missed, but many patients

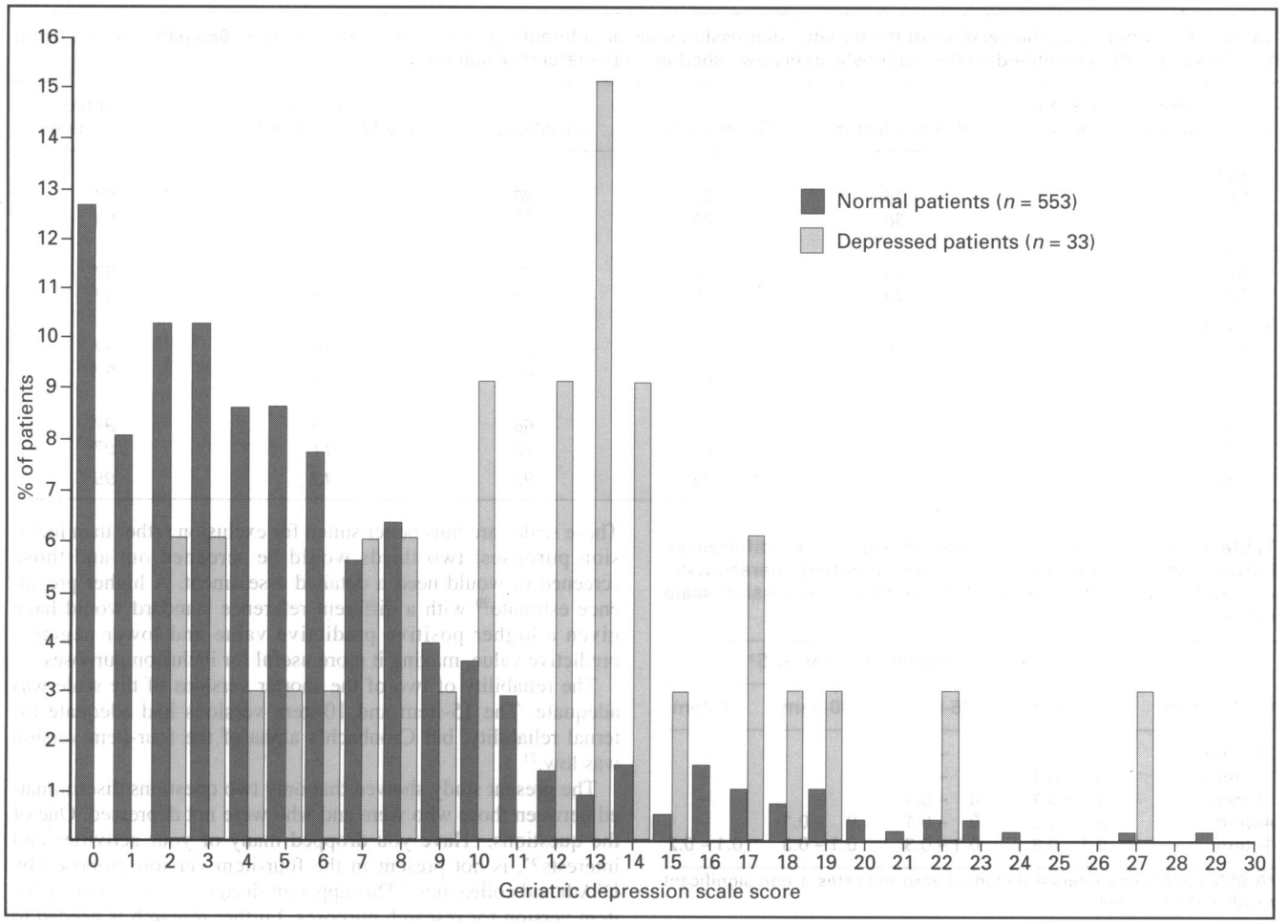

Figure 1. *Frequency of geriatric depression scale scores among normal and depressed patients (*n *= number of patients).*

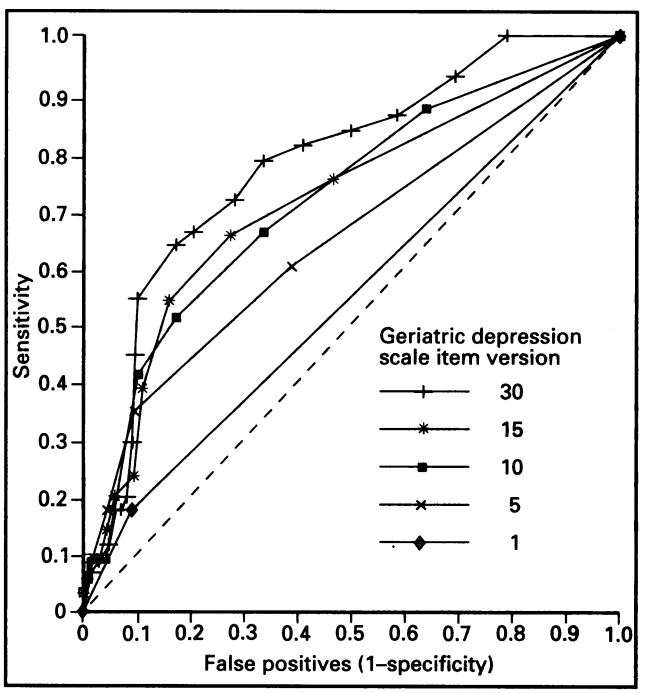

Figure 2. *Receiver operating characteristic curves of the 30-, 15-, 10-, four- and one-item versions of the geriatric depression scale.*

who are not depressed are falsely considered depressed.

The diagnostic performance of the different versions was studied by examining the areas under the receiver operating characteristic curves. Confidence intervals for differences between areas under the curves are presented in Table 2. The 30-item version had an area under the curve of 0.79, the 15-item version of 0.73, the 10-item version of 0.72, and the four-item version of 0.69. The 95% confidence interval of the area under the curve for the one-item version was 0.45–0.66. This contains 0.5, which represents chance. Apart from the one-item version which performed no better than chance, there was no difference in diagnostic value between any of the versions of the scale.

Discriminating items

Logistic regression analysis revealed that the questions: 'Are you basically satisfied with your life'? and 'Have you dropped many of your activities and interests?' were most discriminating between patients who were depressed and not depressed according to the reference interviews ($P<0.05$). None of the other questions added significantly to the discriminating power of the model.

Reliability

Internal-consistency reliabilities of both the 30-item and the 15-item version were good, with a Cronbach's alpha for the 30-item version of 0.87 and for the 15-item version of 0.76. The 10-item and four-item versions had Cronbach's alphas of 0.70 and 0.64, respectively.

Table 1. Performance of the versions of the geriatric depression scale at different cut-off points, using data from 586 patients, and using the 33 patients (6%) identified by the diagnostic interview schedule as the reference standard.

Geriatric depression scale version and cut-off point	Positive test (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
30-item					
<11/11+	*16*	*55*	*86*	*19*	*97*
<7/7+	*36*	*79*	*67*	*12*	*98*
15-item					
<3/3+	*30*	*67*	*73*	*13*	*97*
<2/2+	*49*	*76*	*53*	*9*	*97*
10-item					
<3/3+	*19*	*52*	*83*	*15*	*97*
<2/2+	*36*	*67*	*66*	*10*	*97*
4-item					
<2/2+	*36*	*67*	*66*	*10*	*97*
<1/1+	*30*	*61*	*72*	*11*	*97*
1-item	*8*	*18*	*92*	*13*	*95*

Table 2. 95% confidence intervals (95% CI) of the differences between the areas under the receiver operating characteristic curves for different versions of the geriatric depression scale (GDS).

GDS version	95% CI of differences for GDS[a]			
	30-item	15-item	10-item	4-item
30-item	–	–	–	–
15-item	0 – 0.1	–	–	–
10-item	0 – 0.1	0 – 0.1	–	–
4-item	0 – 0.2	0 – 0.1	0 – 0.1	–
1-item	0.1 – 0.3	0.1 – 0.3	0.1 – 0.3	0.1 – 0.2

[a]A 95% confidence interval including zero indicates a non-significant result at the 5% level.

Discussion

The sample of patients was drawn from consecutive patients attending general practice, both from practice visits and home visits. This sampling frame might have influenced the estimated prevalence. However, open community studies in North America with the diagnostic interview schedule show similar prevalences of major depression and dysthymia in this age group.[19] Practices differed in the number of elderly people they had on their lists so not all were able to recruit the 80 consecutive patients specified in the study. The age–sex distribution was representative of the population. Nonetheless, a substantial proportion (17%) were excluded. A higher prevalence would improve the positive predictive values of the test.

It was found that some patients needed a lot of help in completing the geriatric depression scale. It appears unrealistic to rely on the ability of elderly patients to fill in this questionnaire themselves. This makes it important to have a shorter questionnaire. As the 30-item version was used, the actual feasibility of any of the shorter versions was not tested. The impression was that the difficulties patients experienced were connected to the wording of the questions and not to the number of items. This needs to be studied in a practice setting.

With the exception of the one-item version which performed no better than chance, there was no difference in diagnostic value between any of the versions of the geriatric depression scale. The 15-, 10-, and four-items thus all provided the same diagnostic information as the 30-item version. Given the relatively low prevalence of depression in the group studied, all shorter versions had a low positive and a high negative predictive value. These scales are thus better suited for exclusion rather than inclusion purposes: two thirds would be screened out and those screened in would need a detailed assessment. A higher prevalence estimate[20] with a different reference standard would have given a higher positive predictive value and lower negative predictive value, making it more useful for inclusion purposes.

The reliability of two of the shorter versions of the scale was adequate. The 15-item and 10-item versions had adequate internal reliability, but Cronbach's alpha of the four-item version was low.[21]

The present study showed that only two questions discriminated between those who were and who were not depressed. One of the questions, 'Have you dropped many of your activities and interests?', is not present in the four-item version proposed by D'Ath and colleagues.[9] This apparent disagreement favours a 10-item version for research purposes. Further research is needed to settle this issue and to show which items are most suitable for a four-item version.

The case for classic population screening of elderly general practice patients is unproven at the moment.[22] However, there may be a place for a more selective strategy of case-finding, using questionnaires such as the short versions of the geriatric depression scale or similar instruments[3] such as the short Zung scale.[23] A major problem with this sort of case-finding is the evaluation of depression symptoms in the presence of dementia. Studies on the utility of the geriatric depression scale in demented patients show conflicting results.[24,25]

The 15-, 10-, and four-item versions of the geriatric depression scale are an adequate substitute for the 30-item version. For practical purposes, the smallest subset of questions would be the most desirable: the four-item version. Although the reliability of the four-item version was somewhat low, the four-item version seems preferable because it could perhaps by incorporated into routine care by the general practitioner.

Appendix 1. The 15-item geriatric depression scale, and those items included in the 10-, four- and one-item version of D'Ath and colleagues[9] and in the present study, selected after regression analysis.

	D'Ath[9]	Present study
Are you basically satisfied with your life?	10,4	10, 4, 1
Have you dropped many of your activities and interests?	10	10, 4
Do you feel that your life is empty?	10, 4, 1	
Do you often get bored?		10
Are you in good spirits most of the time?		10
Are you afraid that something bad is going to happen to you?	10, 4	
Do you feel happy most of the time?	10, 4	10, 4

	D'Ath[9]	Present study
Do you feel helpless?	10	10
Do you prefer to stay at home, rather than going out and doing new things?		10, 4
Do you feel you have more problems with memory than most?	10	
Do you think it is wonderful to be alive?		
Do you feel pretty worthless the way you are now?		10
Do you feel full of energy?	10	10
Do you feel that your situation is hopeless?	10	
Do you think that most people are better off than you are?	10	10

References

1. Wells KB, Stewart A, Hays RD, *et al.* The functioning and well-being of depressed patients. *JAMA* 1989; **262:** 914-919.
2. Jorm AF, Henderson AS, Kay DWK, *et al.* Mortality in relation to dementia, depression and social integration in an elderly community sample. *Int J Geriatr Psychiatry* 1991; **6:** 5-11.
3. Bird AS, Macdonald AJD, Mann AH, Philpott MP. Preliminary experience with the selfcare (D): a self-rating depression questionnaire for use in elderly, non-institutionalized subjects. *Int J Geriatr Psychiatry* 1987; **2:** 31-38.
4. Zung WWK, Richards CB, Short MJ. Self-rating depression scale in an outpatient clinic. *Arch Gen Psychiatry* 1965; **13:** 508-515.
5. Yesavage JA, Brink TL, Rose TL, *et al.* Development and validation of a a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1983; **17:** 37-49.
6. Lobo A, Ventura T, Marco C. Psychiatric morbidity among residents in a home for the elderly in Spain: prevalence of disorder and validity of screening. *Int J Geriatr Psychiatry* 1990; **5:** 83-91.
7. Brink TL, Yesavage JA, Lum O, *et al.* Screening tests for geriatric depression. *Clin Gerontologist* 1982; **1:** 37-43.
8. Kok RM, Heeren ThJ, van Hemert Am. De geriatric depression scale (English summary). *Tijdschr Psychiatr* 1993; **35:** 416-421.
9. D'Ath P, Katona P, Mullan E, *et al.* Screening, detection and management of depression in elderly primary care attenders. 1. The acceptability and performance of the 15-item geriatric depression scale (GDS 15) and the development of short versions. *Fam Pract* 1994; **11:** 260-266.
10. Sheikh JL, Yesavage JA. Geriatric depression scale (GDS): recent evidence and development of a shorter version. In: Brink TL (ed). *Clinical gerontology: a guide to assessment and intervention.* New York, NY: Hawthorn Press, 1986.
11. Williams EI, Wallace P. *Health checks for people aged 75 and over. Occasional paper 59.* London: Royal College of General Practitioners, 1993.
12. Royal College of Physicians and British Geriatric Society. *Standardized assessment scales for elderly people.* London: Royal College of Physicians, 1992.
13. Mullan E, Katona P, D'Ath P, Katona C. Screening, detection and management of depression in elderly primary care attenders. 2. Detection and fitness for treatment, a case record study. *Fam Pract* 1994; **11:** 267-270.
14. Dingemans P, Geerlings P, van Limbeek J, Linszen D. *NIMH diagnostic interview schedule, version III. Dutch manual.* Amsterdam, Netherlands: Psychiatric Centre of the Academic Hospital of the University of Amsterdam, 1989.
15. Wouters L, van Limbeek J, d'Hont F. *The DIS program package.* Amsterdam, Netherlands: Psychiatric Centre of the Academic Hospital of the University of Amsterdam, 1989.
16. Sackett DL, Hayes RB, Tugwell P. *Clinical epidemiology. A basic science for clinical medicine.* Boston, MA: Little, Brown and Company, 1985.
17. Hanley JA, McNeill BJ. The meaning and use of the area under a receiver operating curve (ROC). *Radiology* 1982; **143:** 29-36.
18. Hanley JA, McNeil BJ. A method of comparing the areas under receiving operating characteristic curves derived from the same cases. *Radiology* 1983; **148:** 839-843.
19. Bland RC, Newman SC, Orn H. Prevalence of psychiatric disorders in the elderly in Edmonton. *Acta Psychiatr Scand* 1988; **77** (suppl 338): 57-63.
20. Livingston G, Hawkins A, Graham N, *et al.* The Gospel Oak study: prevalence rates of dementia, depression and activity limitation among elderly residents in inner London. *Psychol Med* 1990; **20:** 137-146.
21. Zyzanski SJ. Cutting and pasting new measures from old. In: Stewart M, Tudiver F, Bass MJ, *et al* (eds). *Tools for primary care research.* Newbury Park, CA: Sage, 1992.
22. Wright AF. Should general practitioners be testing for depression? *Br J Gen Pract* 1994; **44:** 132-135.
23. Tucker MA, Ogle SJ, Davison JG, Eilenberg MD. Validation of a brief screening test for depression in the elderly. *Age Ageing* 1987; **16:** 139-144.
24. O'Riordan TG, Hayes JP, O'Neill D, *et al.* The effect of mild to moderate dementia on the geriatric depression scale and on the general health questionnaire. *Age Ageing* 1990; **19:** 57-61.
25. O'Neil D, Rice I, Blake P, *et al.* The geriatric depression scale: rater-administered or self-administered? *Int J Geriatr Psychiatry* 1992; **7:** 511-515.

Acknowledgements

A grant for this study was awarded by the Dutch Organization for Scientific Research (NWO). We thank Professor Cornelius Katona for the use of data.

Address for correspondence

Dr H W J van Marwijk, Department of General Practice, Leiden University, PO Box 2088, 2301 CB Leiden, Netherlands.

Food for thought...

'Counselling is a social act, not a chemical behaviour. Politics and economics are also social sciences, yet often those very market theorists who exalt judgement by outcome are content to impose ideologically-driven changes in advance of empirical testing'.

Hazzard AJ. Measuring outcome in counselling: a brief exploration of the issues [editorial].
March *Journal*, p. 118.