BRIEF COMMUNICATION

Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in the Netherlands

A. T. F. BEEKMAN,[1] D. J. H. DEEG, J. VAN LIMBEEK, A. W. BRAAM, M. Z. DE VRIES AND W. VAN TILBURG

From the Department of Psychiatry of the Free University Amsterdam, The Netherlands

SUMMARY   The Center for Epidemiologic Studies Depression scale (CES-D) has been widely used in studies of late-life depression. Psychometric properties are generally favourable, but data on the criterion validity of the CES-D in elderly community-based samples are lacking. In a sample of older (55–85 years) inhabitants of the Netherlands, 487 subjects were selected to study criterion validity of the CES-D. Using the 1-month prevalence of major depression derived from the Diagnostic Interview Schedule (DIS) as criterion, the weighted sensitivity of the CES-D was 100%; specificity 88%; and positive predictive value 13·2%. False positives were not more likely among elderly with physical illness, cognitive decline or anxiety. We conclude that the criterion validity of the CES-D for major depression was very satisfactory in this sample of older adults.

INTRODUCTION

Depression is one of the most common disorders of old age (Blazer, 1989). It is associated with declines in quality of life (Gurland, 1992) and functioning (Ormel *et al*. 1993, 1994). Although depression is generally thought to be a treatable disorder, only a minority of depressed elderly people receive adequate treatment (Copeland *et al*. 1992). Recognition and diagnosis of depression in primary care is complicated and has been shown to be rather poor (Goldberg & Bridges 1988; Copeland *et al*. 1992). In the past decades a number of screening tests were developed to improve recognition of depression. These tests are also used as first-stage screening devices in epidemiological studies. The Center for Epidemiologic Studies Depression scale (CES-D, Radloff, 1977) has been used in many community based studies in the elderly (Radloff & Teri 1986, Beekman *et al*. 1994, 1995). However, research regarding the criterion validity of the CES-D has been limited to younger samples. In these studies the sensitivity of the

CES-D for major depression varied between 60 and 99%, and its specificity between 73 and 94% (Beekman *et al*. 1994). Total scores range from 0–60. Although the cut-off for a depressive syndrome which is probably clinically relevant is generally 16, some authors claim this too low (Husaini *et al*. 1980; Himmelfarb & Murrell, 1983). Clearly, data on the criterion validity of the CES-D in older samples are necessary before conclusions regarding its usefulness as a screenings instrument can be drawn. In this paper we report on the criterion validity of the CES-D in a community-based sample of older adults in the Netherlands. Special attention was paid to co-morbid anxiety, cognitive dysfunctioning and physical disease, because these variables are known to complicate diagnosis of depression in the elderly (Reifler, 1994).

METHOD

Sampling and procedures

The Longitudinal Aging Study Amsterdam (LASA) is a 10-year longitudinal study (Deeg *et al*. 1993). Full details on sampling and response are described elsewhere (Beekman *et al*. 1995). In short, a random sample of older (55–85)

[1] Address for correspondence: Dr Aartjan T. F. Beekman, Department of Psychiatry, Valerius Clinic, Valeriusplein 9, 1075 BG Amsterdam, The Netherlands.

persons, stratified for age, sex and level of urbanization was drawn from the population registers in 11 municipalities in the Netherlands. The sample was used in two studies. Respondents were first interviewed for the NESTOR program Living arrangements and Social Networks of older adults (response 62·3%, Broese *et al.* 1994). About 10 months later, 3107 (81·7%) of the 3805 respondents to the NESTOR-LSN study took part in the LASA baseline interview. Non-response was related to age ($P < 0.001$), but not to sex. The older old were more often found to be too ill or cognitively impaired to participate. Due to item non-response a further 51 subjects were lost, leaving a baseline sample of 3056.

The baseline interview included the CES-D and all instruments except the Diagnostic Interview Schedule (DIS, Robins *et al.* 1981), which was used as diagnostic criterium. The DIS was included in a medical interview (response 86·0% relative to baseline) in all subjects scoring $\geqslant 16$ on the CES-D and in a random sample of those scoring < 16. Again, non-response was related to age ($P < 0.001$), but not to sex. For this paper, respondents were further selected on the basis of the lag-time between baseline and medical interview. The CES-D has a time-range of 1 week. In the DIS depressive symptoms have to be present for at least 2 weeks before they are considered in diagnosis. Taking the 1-month recency of DIS diagnoses as a base, the maximum lag-time allowable is 6 weeks. For this paper then, only subjects with lag-times less than 42 days were included. This reduced the study population to 487 (average lag-time 20·7 days). Interviews were conducted by especially trained interviewers in the homes of respondents between September 1992 and October 1993. Screening and diagnostic interviews were administered by separate interviewers.

Measures

The DIS is a criterion instrument designed for epidemiological studies (Robins *et al.* 1991). We used an adapted version, including only the sections on DSM-III affective and anxiety disorders. Two main points of critique on the DIS have been that life-time diagnoses are unreliable (Knauper & Wittchen, 1994) and that prevalence rates of depression in elderly subjects are too low (Snowdon, 1990; Heithoff, 1995).

Prevalence rates may be low because depressive symptoms, which according to the subject are caused by physical illness, or use of medication, alcohol or drugs, are rigorously ignored in diagnosis. In this paper, life-time diagnoses are used only for dysthymic disorder. In order to study effects of ill health on the relationship between the CES-D and the DIS, respondents were asked detailed questions on chronic diseases (CBS, 1989) and functional limitations (van Sonsbeek, 1988). The Mini-Mental State Examination (MMSE, Folstein *et al.* 1975) was used to indicate cognitive functioning. Symptom scores of anxiety were measured using the anxiety subscale of the Hospital Anxiety and Depression scale (HADS-A, Zigmond & Snaith, 1983). This scale was chosen because it represents a measure of anxiety unrelated to depression. Finally, five basic demographic variables were included (age, sex, marital status, level of education and living arrangement).

Statistics

Parameters for the criterion validity of the CES-D calculated include positive and negative predictive values (Bouter & van Dongen, 1995); and sensitivity and specificity (Bland, 1990). Using the standard deviation of a binomial distribution (Bland, 1990), 95% confidence intervals were calculated. Characteristics of true and false positives were compared using χ^2 statistics and logistic regression.

RESULTS

In Table 1 characteristics of the sample are shown. As intended, about half the sample was depressed according to the CES-D. One-fifth had elevated anxiety scores. Men and women were roughly evenly represented. Two-thirds of the sample had one or more chronic physical disease, and about half had one or more functional limitation. Sixteen per cent scored less than 24 points on the MMSE. This shows that attrition has not caused the sample to become a sample of 'healthy elderly'. There was only one institutionalized subject.

In Table 2 sensitivity and specificity of the CES-D are shown. With regard to major depression, high sensitivity and low specificity may indicate that the cut-off used on the CES-D was too low. Analyses were, therefore, repeated

Table 1. *Characteristics of the study-sample*

Variable	N	(%)
Age (years)		
55–64	154	(32)
65–74	138	(28)
75–85	195	(40)
Sex		
Male	205	(42)
Female	282	(58)
Level of education		
Low	218	(45)
Middle	204	(42)
High	65	(13)
Marital status		
Married	266	(55)
Not/no longer married	221	(45)
Living arrangement		
Independent	465	(96)
Old age residence	21	(4)
Nursing home/hospital	1	(0)
Total CES-D score		
< 16	252	(52)
⩾ 16	235	(48)
Total HADS anxiety score*		
< 7	363	(79)
⩾ 8	97	(21)
Mini-Mental State†		
24–30	400	(83)
0–23	84	(17)
Number of chronic diseases‡		
None	171	(35)
One	167	(34)
Two or more	148	(31)
Functional limitations†		
None	245	(51)
One or more	237	(49)

* 27 missing observations.
† 3 missing observations.
‡ 1 missing observation.
† 5 missing observations.

$85 \cdot 1 : 14 \cdot 9 = 5 \cdot 7 : 1$), was necessary to extrapolate the findings back to baseline (Table 3). This causes sensitivity to drop from 100 % for recent episodes of major depression, to 70 % for episodes in the past year. For dysthymic and anxiety disorders sensitivity is very poor at approximately 40 %. Specificity is approximately 88 % for all disorders studied. Re-examining the effect of changing the cut-off, sensitivity for recently experienced episodes of major depression now drops from 100 % with a cut-off of ⩾ 16 to 72 % with a cut-off ⩾ 20. Positive and negative predictive values remain unchanged by weighting.

The low positive predictive value indicates that the majority of those scoring above the threshold on the CES-D did not reach the DSM-III criteria for major depression. False positives were less likely to be female ($P < 0.05$); more likely to be older-old (i.e. aged > 75 years) ($P < 0.01$); and equally likely to have received

Table 2. *Percentages for sensitivity and specificity of the CES-D*

Disorder (recency)	CES-D score ⩾ 16		CES-D score ⩾ 18		CES-D score ⩾ 20	
	Sens	Spec	Sens	Spec	Sens	Spec
Major depression (past month)	100	55·3	93·5	65·6	93·5	73·5
Major depression (past year)	93·2	56·2	86·4	66·6	86·4	74·7
Dysthymic disorder (life-time)	75·0	54·7	64·6	64·7	64·6	72·9
All anxiety disorders (past year)	74·1	57·2	68·2	68·2	61·2	75·6

Table 3. *Percentages for weighted sensitivity and specificity of the CES-D**

Disorder (recency)	CES-D score ⩾ 16		CES-D score ⩾ 18		CES-D score ⩾ 20	
	Sens	Spec	Sens	Spec	Sens	Spec
Major depression (past month)	100	87·6 (86–89)	71·8 (58–86)	91·4	71·8 (58–86)	94·0
Major depression (past year)	70·6 (59–82)	88·0	52·6 (41–64)	91·9	52·6 (41–64)	94·4
Dysthymic disorder (life-time)	40·8 (31–50)	86·7	24·2 (17–31)	91·3	24·2 (17–31)	93·9
All anxiety disorders (past year)	40·3 (33–47)	88·1	27·4 (21–33)	92·5	21·7 (17–27)	94·6

* 95% confidence intervals shown in parentheses. Because the intervals are very similar across specificities, only one confidence interval is reported here.

using 18 and 20 as the cut-off. The drop in sensitivity is small compared to the gains made in specificity. Because of the low prevalence of major depression in the source population (one-month prevalence = 2 %, Beekman *et al.* 1995), positive predictive values were low, ranging from 13·2 % with the cut-off at ⩾ 16, to 19·3 % at ⩾ 20. Negative predictive values were very high at 99–100 %. In the baseline sample 14·9 % scored above the cut-off (⩾ 16) on the CES-D (Beekman *et al.* 1995). In the study sample this was stratified to be 50 %. As stratification influences sensitivity and specificity rather heavily, recalculation, using appropriate weights (subjects scoring < 16: those scoring ⩾ 16 =

higher education or to be married. True and false positives were similar with regard to physical health and cognitive performance. High levels of anxiety were present in 40 % of the false positives and 78 % of the true positives ($P < 0.001$). In multiple logistic regression the association with age was no longer significant.

DISCUSSION

To our knowledge this is the first report of the criterion validity of the CES-D in a large community-based sample of older subjects. Stratification for age and sex facilitates studying the validity of the CES-D in men and women throughout older age, but, as non-response was related to age, also has lead to lower response. For the present purpose it is important to note that factors that are notorious for complicating the diagnosis of depression in the elderly (physical illness, cognitive decline and anxiety), were all well represented. Although non-response was higher among the older-old, this has not resulted in a sample of 'healthy elderly'. Subjects living in institutions were clearly under-represented.

Considering the results, the sensitivity of the CES-D for major depression (1-month prevalence) was excellent. For use in the community, extrapolation to the source sample gives the most relevant results. Using 16 as the cut-off, sensitivity for major depression was 100 %, while specificity was 88 %. Using a higher cut-off improved the specificity at the expense of the sensitivity. As allowing sensitivity to drop below 80 % is incompatible with most screening purposes, it seems advisable to use the traditional cut-off ($\geqslant 16$). Considering the extrapolated results, it appears that the sensitivity of the CES-D was satisfactory for major depression only. The sensitivity for dysthymic and anxiety disorders was much lower.

The majority of those depressed according to the CES-D did not fulfil rigorous diagnostic criteria for DSM-III affective disorders (positive predictive value 13·2 %). True and false positives were equally likely to be physically ill or cognitively impaired, while anxiety levels were highest among those with major depression.

We conclude that the CES-D has satisfactory criterion validity for major depression in this sample of the older population in the Nether-

lands. Further study is however necessary to determine the validity of the CES-D in other settings, such as general hospitals, or nursing homes.

REFERENCES

Beekman, A. T. F., Limbeek, J. v., Deeg, D. J. H., Wouters, L. & Tilburg, W. van (1994). Screening for depression in the elderly in the community: using the Center for Epidemiologic Studies Depression scale (CES-D) in the Netherlands. *Tijdschrift voor Gerontologie en Geriatrie* **25**, 95–103.

Beekman, A. T. F., Deeg, D. J. H., Tilburg, T. v., Smit, J. H., Hooijer, C. & Tilburg, W. van (1995). Major and minor depression in later life: a study of prevalence and risk factors. *Journal of Affective Disorders* **36** 65–75.

Bland, M. (1990). *An Introduction to Medical Statistics*. Oxford Medical Publications: Oxford.

Blazer, D. G. (1989). Depression in the elderly. *New England Journal of Medicine* **320**. 164–166.

Bouter, L. M. & Dongen, M. C. J. M. (1995). *Epidemiologisch onderzoek: opzet en interpretatie*, pp. 291–297. Bohn Stafleu: Houten.

Broese van Groenou, M. I., Tilburg, T. G. v., Leeuw, E. D. de & Liefbroer, A. C. (1994). Data collection. In *Living Arrangement and Social Networks of Older Adults, First Results* (ed. C. P. M. Knipscheer, J. de Jong-Gierveld, T. G. van Tilburg, P. A. Dijkstra), pp. 1–19. Vrije Universiteit: Amsterdam.

Central Bureau of Statistics (CBS), (1989). *Health Interview Questionnaire*. Heerlen: The Netherlands.

Copeland, J. R. M., Davidson, I. A., Dewey, M. E., Gilmore, C., Larkin, B. A., McWilliam, C., Saunders, P. A., Scott, A., Sharma, V. & Sullivan, C. (1992). Alzheimer's disease, other dementias, depression and pseudodementia: prevalence, incidence and three-year outcome in Liverpool. *British Journal of Psychiatry* **161**, 230–239.

Deeg, D. J. H., Knipscheer, C. P. M., Tilburg, W. van (1993). Autonomy and well-being in the aging population: concepts and design of the Longitudinal Aging Study Amsterdam. *NIG Trend-studies No. 7*. Netherlands Institute of Gerontology: Bunnik.

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.

Goldberg, D. & Bridges, K. (1988). Detecting anxiety and depression in general medical settings. *British Medical Journal* **297**, 897–899.

Gurland, B. J. (1992). The impact of depression on quality of life of the elderly. *Clinical Geriatric Medicine* **8**, 377–386.

Heithoff, K. (1995). Does the ECA underestimate the prevalence of late-life depression. *Journal of the American Geriatric Society* **43**, 2–6.

Himmelfarb, S. & Murrell, S. A. (1983). Reliability and validity of five mental health scales in older persons. *Journal of Gerontology* **38**, 333–339.

Husaini, B. A., Neff, J. A., Harrington, J. B., Hughes, M. D. & Stone, R. H. (1980). Depression in rural communities: validating the CES-D scale. *Journal of Community Psychology* **8**, 20–27.

Knauper, B. & Wittchen, H. U. (1994). Diagnosing major depression in the elderly: evidence for response bias in standardised diagnostic interviews. *Journal of Psychiatric Research* **28**, 147–164.

Ormel, J., VonKorff, M., Van den Brink, W., Katon, W., Brilman, E. & Oldehinkel, T. (1993). Depression, anxiety, and social disability show synchrony of change in primary care patients. *American Journal of Public Health* **83**, 385–390.

Ormel, J., VonKorff, M., Ustun, T. B., Pini, S., Korten, A. & Oldehinkel, T. (1994). Common mental disorders and disability across cultures. *Journal of the American Medical Association* **272**, 1741–1748.

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* **1**, 385–401.

Radloff, L. S. & Teri, L. (1986). Use of the CES-D with older adults. *Clinical Gerontologist* **5**, 119–136.

Reifler, B. V. (1994). Depression: diagnosis and comorbidity. In *Diagnosis and Treatment of Depression in Late Life. Results of the NIH Consensus Conference*, pp. 55–61. American Psychiatric Press: Washington.

Robins, L., Helzer, J. E., Croughan, J. & Radcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. *Archives of General Psychiatry* **38**, 381–389.

Snowdon, J. (1990). The prevalence of depression in old age. *International Journal of Geriatric Psychiatry* **5**, 141–144.

Sonsbeek, J. L. A. v. (1988). Methodological and substantial aspects of the OECD indicator of chronic functional limitations. *Maandbericht Gezondheid (CBS)* **88**, 4–17.

Zigmond, A. S. & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica* **67**, 361–370.