Multiple assessments of depressive symptoms as an index of depression in population-based samples

Nyklicek, I.; Scherders, M.J.; Pop, V.J.M.

ELSEVIER

Multiple assessments of depressive symptoms as an index of depression in population-based samples

Ivan Nyklíček [a,b,*], Mark J. Scherders [a,c], Victor J. Pop [a,b]

[a] *Research Unit of the Diagnostic Center Eindhoven, Eindhoven, The Netherlands*
[b] *Department of Psychology and Health, Tilburg University, Tilburg, The Netherlands*
[c] *Department of Psychiatry, Catharina Hospital, Eindhoven, The Netherlands*

Abstract

The incremental validity of repeated measurements of a short depression questionnaire was examined regarding the clinical diagnosis of depression. Participants were 951 randomly selected women of around menopausal age. They completed the Edinburgh Depression Scale (EDS) at two time points, with approximately 18 months in between. At the second time point, they participated in a structured clinical interview for depression diagnosis based on the Research Diagnostic Criteria (RDC). With repeated assessments, specificity and negative predictive value (NPV) did not change much relative to a single assessment, with a specificity of 95.0% and a NPV of 91.7% at a cut-off score of 12 on the EDS. As expected, sensitivity dropped, from 87.9% to 58.8%. However, positive predictive value (PPV) increased from 42.0% to 49.1% at a cut-off of 12. When using a cut-off score of 15 on the EDS, the PPV based on both EDS measurements reached 61.8%, yielding a 25-fold probability of being a case for women scoring above 15 at both time points (OR = 24.54, 95% CI = 14.24–42.28). In conclusion, the 10-item EDS is a reliable, valid and valuable screening instrument. When employed repeatedly, a more stable depression may be tapped, which can be of substantial value for both epidemiological research and clinical practice.
© 2004 Elsevier Ireland Ltd. All rights reserved.

Keywords: Assessment; Depression; Incremental validity; Predictive validity

1. Introduction

Depression is a widespread and often chronic affective disorder (World Health Organization, 1998). In addition, a large proportion of depressed individuals are not diagnosed as such, but they frequently attend somatic health services, which contrib-

utes to depression placing a heavy financial burden on society (Lave et al., 1998; Panzarino, 1998). Depression is not only a powerful impediment to psychological well-being, but evidence is accumulating that it is an important risk factor for physical disease as well (Penninx et al., 1999; Stoltz et al., 1999). For instance, an increasing number of studies have shown that depression is a predictor of all-cause mortality, and also of more specific health problems, such as cardiovascular disease, (auto-)immune disorders, low bone mineral density and potentially even cancer (Zonderman et al., 1989; Schweiger et al., 1994;

* Corresponding author. Department of Psychology and Health, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands. Tel.: +31-13-4662391; fax: +31-13-4662370.

E-mail address: i.nyklicek@uvt.nl (I. Nyklíček).

Leonard and Miller, 1995; Glassman and Shapiro, 1998; Musselman et al., 1998; Penninx et al., 1998; Pop et al., 1998). Given these associations, the issue of depression has become progressively more important in epidemiological studies on predictors of physical health outcomes (Glassman and Shapiro, 1998; Musselman et al., 1998; Nyklíček et al., 2003). In such studies, it is crucial to have a large number of respondents to conduct analyses with sufficient power. Given the time-consuming and costly nature of conducting a psychiatric interview to assess the presence of depression, most researchers feel obliged to turn to questionnaires to assess depression or, rather, depressive symptomatology as a proxy for depression (Musselman et al., 1998; Penninx et al., 1998).

It is of utmost importance to scrutinize the relation between scores on questionnaires assessing depressive symptomatology and clinical depression diagnosis, as established on the basis of a diagnostic interview. Several cross-sectional studies on this topic have shown adequate validity of questionnaire measures, indicating that a high score on such a self-rating scale is the strongest predictor of the presence of a depression diagnosis (Zich et al., 1990; Becht et al., 2001). However, since repeated assessments of depressive symptoms may predict depression better than a single determination, it would be of substantial relevance to establish the additional value of multiple questionnaire assessments over a single measurement. Moreover, knowing the predictive power of multiple questionnaire assessments for the presence of a depression diagnosis would be of significance for epidemiological investigations using depressive symptoms as predictors of health, as well as for depression assessment in clinical practice (Zich et al., 1990). Regarding the former issue, an example showing the potential importance of repeated assessment of depressive symptoms is a recent study in which it was demonstrated that while a single measurement of a depressive state did not predict risk for cancer, multiple measurements did show an enhanced risk of future cancer development in persons scoring consistently high on a measure of depressive symptoms (Penninx et al., 1998).

Therefore, in the present study, the associations were examined between repeated scores on a short questionnaire assessing depressive symptomatology, with on average an 18-month interval between the measurements, and diagnosis of depression based on a clinical interview.

2. Methods

2.1. Respondents

Between September 1994 and October 1995 (T1), all women born between 1941 and 1947 ($N = 8503$), living in Eindhoven, The Netherlands, were invited to take part in a large screening study on the prevalence of osteoporosis in perimenopausal women: the Eindhoven Perimenopausal Osteoporosis Study (EPOS; Smeets-Goevaers et al., 1998). A total of 6846 (81%) agreed to participate. Of these women, 950 women belonging to ethnic minorities were excluded because of potential language problems. The remaining 5896 women were asked to complete a set of questionnaires, including a depression self-rating scale, at home and to return it within 1 week. Complete data were obtained from 4944 (83.8%). This group did not differ in baseline characteristics from the original study population. More detail regarding this group has been reported elsewhere (Smeets-Goevaers et al., 1998; see the flow-chart in Fig. 1).

Of the women who agreed to participate in the screening, 1510 (22%) were randomly selected for a follow-up 90-min interview at home, to assess the presence of syndromal depression. Eighty-two percent of this population ($N = 1242$) agreed to participate in this interview, which on average took place
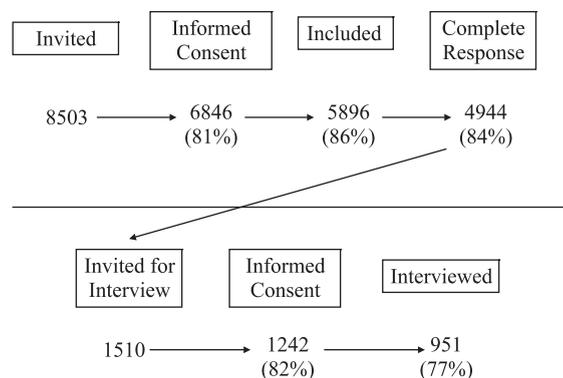

Fig. 1. Flow chart of the study.

18.4 months (S.D. = 4.1 months) later (T2). Besides the assessment of clinical depression, participants again completed the questionnaire assessing depressive symptomatology. Incomplete data and not being Dutch Caucasian resulted in the exclusion of 150 and 141 women, respectively. This resulted in a final sample of 951 women with complete interview data. Of these women, aged between 47 and 55 years, 78% were married or living together, 12.5% were divorced, 5.8% single and 3.3% widowed. Twenty-one women were currently being treated for depression, while 287 (30%) had ever received a treatment for depression.

2.2. Instruments

2.2.1. Depressive symptomatology

Depressive symptoms were assessed using a 10-item self-rating scale initially called the Edinburgh Postnatal Depression Scale (EPDS; Cox et al., 1987), since the original purpose was to assess depressive symptoms in postpartum women. The scale has also been validated in non-childbearing women (Cox et al., 1996), resulting in a new nomenclature: the Edinburgh Depression Scale (EDS). The Dutch version of the EDS has been validated both in women during the postpartum period (Pop et al., 1992) and recently in menopausal women (Becht et al., 2001). Items are scored on four-point rating scales. Total scores can range between 0 and 30, with cut-off scores usually between 11 and 13 (Cox et al., 1987; Harris et al., 1989; Murray and Carothers, 1990).

2.2.2. Clinical depression

Clinical diagnosis of depression was made using the structured diagnostic interview method according to Research Diagnostic Criteria (RDC; Spitzer et al., 1978). The interview was performed by health care nurses, who received special 5-day training to perform the diagnostic interview. These interviewers were unaware of the results of the questionnaires.

2.3. Statistical analysis

Statistical analyses were performed using SPSS software. The predictive power of the EDS (at T1, T2, and at T1 and T2 taken together) with respect to the presence of a major depression diagnosis at T2 was tested by computing sensitivity (the percentage of cases correctly identified by the questionnaire), specificity (the percentage of non-cases correctly identified), positive predictive value (PPV; the percentage of high scorers on the EDS who were cases at interview) and negative predictive value (NPV; the percentage of EDS low scorers who indeed appeared to be non-cases at interview). Since (i) the ideal sensitivity/specificity balance varies with the different clinical and epidemiological purposes; (ii) this balance varies across different cut-off points; and (iii) in the literature no agreement exists as to which cut-off score should be used (Cox et al., 1987; Harris et al., 1989; Murray and Carothers, 1990; Becht et al., 2001), for each time point the EDS scores were dichotomized systematically along cut-off points ranging from 11 to 15. For classification based on multiple time points, respondents that scored consistently above or below the cut-off point were compared with the rest of the sample.

3. Results

The mean EDS score of the 956 interviewees with complete data was 7.27 (S.D. = 6.04), which was not different from the mean EDS score of the original EPOS sample (also $M = 7.27$, S.D. = 6.04, t (5564) = 0.01, $P > 0.50$). Depression was diagnosed in 21.8% of the 956 interviewees with complete data, of whom 11.4% had minor depression and 10.4% had major depression. When using the frequently applied cut-off point of 12 on the EDS (Cox et al., 1996), 22.2% of the respondents could be regarded as being depressed at the same time (T2). At T1, using the same cut-off score, 23.3% of the 5566 respondents were rated as being depressed. Respondents who indicated having ever been treated for depression ($N = 287$) had substantially higher EDS scores at both T1 ($M = 10.71$, S.D. = 6.55) and T2 ($M = 11.14$, S.D. = 6.65) than never-treated participants ($M = 5.51$, S.D. = 5.03 and $M = 6.04$, S.D. = 4.87, respectively) ($t > 12.0$, $P < 0.001$). In addition, 66 individuals (23.0%) who had ever been treated for depression received a diagnosis of major depression at the present interview, compared with only 33 (4.9%) of the never-treated women ($\chi^2 = 70.59$, $P < 0.001$). These figures did not change substantially when the 21 women treated at present were excluded from the analyses. The corre-

lation between the EDS scores at both time points was $r = 0.65$ ($P < 0.001$), indicating a reasonably high stability of depressive symptoms across an approximately 18-month period.

The sensitivity, specificity, PPV and NPV of the EDS depression classifications at T1, T2, and both T1 and T2 taken together, with respect to the RDC diagnosis of major depression can be viewed in Table 1. All κ values, as a measure of association, were highly significant ($P < 0.001$); the lowest κ was found for cut-off point 11 at T1 (0.278, $t = 9.57$, $P < 0.001$), at T2 the lowest κ was 0.429 ($t = 15.23$, $P < 0.001$), and for the analyses based on both time points combined, the lowest κ was 0.451 ($t = 13.74$, $P < 0.001$). As can be seen, the coefficients of the EDS scores at T2 are higher than those at T1, with sensitivity at T2 ranging from 72.7% at the cut-off point of 15 to 89.9% at the cut-off point of 11, specificity ranging between 87.0% (11) and 95.9% (15), PPV ranging from 36.0% (11) to 55.8% (15), and NPV between 87.0% (15) and 92.2% (11). All values were substantially lower at T1, compared with T2, specificity and NPV only slightly, but sensitivity and PPV substantially. Sensitivity had values between 53.6% at a cut-off of 15 and 67.0% (cut-off 11), while PPV ranged from 27.4% (at cut-off 11) to 39.4% (cut-off 15).

Classification based upon EDS scores at both T1 and T2 slightly improved the specificity and NPV, and moderately PPV, but sensitivity dropped substantially to a level below that found at T1. The percentages of specificity were approximately 2–5 points higher (reaching 98.0% at cut-off 15), while PPV increased by some 6–8 points, reaching 61.8% at cut-off 15. This means that approximately 6 out of 10 persons scoring at least 15 on the EDS on both occasions would be classified as having major depression according to the RDC. These women had a 25-fold risk of being diagnosed with major depression, compared with

women not scoring at least 15 on both occasions (odds ratio = 24.54, 95% CI = 14.24–42.28).

4. Discussion

When using the widely used cut-off score of 12 on the EDS, between 22.2% (T1) and 26.5% (T3) of the respondents may be regarded as being depressed in our sample of women around menopausal age. This corresponds closely to outcomes obtained earlier in non-childbearing women, using the EDS (27.2%; Cox et al., 1996), and in middle-aged women, using the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977): 26% (Kaufert et al., 1992). Furthermore, the percentages found are somewhat higher than the percentage of women who were diagnosed as having either minor or major depression, according to the RDC in the present study (21.8%). In addition, major depression was diagnosed in 10.4% of the respondents, which is very similar to the rate of 11.3% cases of diagnosed depression in a recent large Dutch epidemiological study (NEMESIS, $N = 7076$; Bijl et al., 1997), in which the CIDI was used for assessing depression.

According to Table 1 and as reported earlier (Becht et al., 2001), the EDS appropriately detects major depression, as assessed by the RDC, with sensitivities ranging from 72.7% to 89.9%, depending on the cut-off criterion. The PPVs are lower, but still substantial: between 36% (at cut-off 11) and 55.8% (at cut-off 15). The specificity and NPV are quite high (high eighties to lower nineties). Although PPV values were adequate, in order to examine possible concomitants of classification error rates, false positives (those having high EDS scores at T2, but a negative interview result) and women correctly classified as depressed were compared on a number of characteristics, such as

Table 1
Validity of the Edinburgh Depression Scale at T1, T2, and at T1 and T2 together (T12)

	Sensitivity (%)			Specificity (%)			PPV (%)			NPV (%)		
Cut-off	T1	T2	T12	T1	T2	T12	T1	T2	T12	T1	T2	T12
11	67.0	89.9	64.9	82.0	87.0	92.2	27.4	36.0	43.8	86.9	92.2	92.8
12	61.9	87.9	58.8	85.3	90.2	95.0	29.1	42.0	49.1	86.3	90.5	91.7
13	59.8	83.8	55.7	88.5	92.4	96.2	32.4	46.4	52.4	86.2	89.3	90.6
14	58.8	78.8	54.6	90.7	94.9	97.3	35.2	52.7	58.9	86.2	88.1	89.9
15	53.6	72.7	48.5	93.4	95.9	98.0	39.4	55.8	61.8	85.3	87.0	88.8

marital status, menopausal status, and EDS scores at T1. No significant differences between these groups were found, except for a tendency for the high-EDS-only women to be post-menopausal ($P = 0.073$). This may suggest that a part of the elevated EDS scores may be a consequence of post-menopausal status rather than of depression per se. However, since the EDS does not contain items reflecting physical symptoms, this explanation does not seem very likely. Together with the fact that this effect was only marginally significant, and may therefore be due to chance, it is concluded that no determinants of the present reasonably low misclassification rates have been identified.

When scores on the EDS obtained approximately 18 months before the interview are used, sensitivity is substantially lower compared with scores obtained at the same time as the diagnostic interview. This finding is obtained despite the fairly high correlation between the scores on the EDS at both time points (0.65), which indicates a reasonable overall stability of EDS scores over this period. These at first sight seemingly contradictory findings are in accordance with studies showing that measures of negative affect have both a strong stable component, as well as a smaller, but substantial component that is subject to environmental influences (Ormel and Schaufeli, 1991). Finally, the PPVs at T1 are clearly lower than the PPVs obtained at T2, but they still reflect a moderate predictive power of the EDS regarding RDC diagnosis of depression approximately 18 months later (up to 39.4% at cut-off score 15).

The main purpose of the present study, however, was to examine to what extent repeated high EDS scores would tap a stable aspect of depression and therefore be a better predictor of a depression diagnosis. Specificity and NPV improved only slightly, while sensitivity dropped substantially, to a level lower than that based on the first measurement, 18 months before the interview. This means that the probability that a case will score high on both depressive symptoms measurement occasions is substantially lower than the probability of having elevated scores at a single measurement performed simultaneously with the interview. This is not surprising, given the strong probability that a part of the major depressives developed the condition between the two measurements. However, PPV, which is the most relevant measure when predictive power is concerned, was enhanced by some 6–8 points, reaching 61.8% at the cut-off score 15. This means that 6 out of 10 women scoring 15 or more on the EDS on both occasions were diagnosed as having major depression (for both minor and major depression together, this figure would be 82.9%, an increase of 6.2), a 53-fold risk compared with women who scored below 15 on both occasions.

These results indicate good psychometric properties of the EDS in the present population of women around the menopause. Moreover, these results show a clear additional value of repeated measurements of the EDS with respect to the prediction of the presence of major depression. For epidemiological research purposes, these results imply that the EDS can be used as a proxy for the presence of a depressive state. In addition, when this scale is applied two times with approximately 18 months between the measurements, the reasonable PPV of about 60% (at cut-off points 14 and 15) suggests that in this way the EDS may even tap an important part of a more stable, chronic, depression (Penninx et al., 1998). This would be ideal for epidemiological research into, for instance, enhanced risk for diseases such as cardiovascular disease or cancer. Interestingly, as discussed in Section 1, in a recent study the additional predictive value of multiple measurements of depressive symptoms over a single assessment has been clearly demonstrated regarding future cancer development (Penninx et al., 1998).

For clinicians too, these results may be of relevance, especially clinicians working in primary health care. Given the facts that (a) costs of health care services have become a heavy burden for Western societies (Feldman, 2000); (b) clinicians have less time to spend on a patient; and (c) increasing numbers of patients visit primary health care centers with vague symptoms (Holder-Perkins et al., 2000), a fast, cost-effective first screening for depression would be a welcome aid.

We may conclude that the EDS is a reliable, valid, and valuable instrument for both epidemiological screening studies, especially in the case of repeated administration, and in clinical practice as a first screening on which basis a decision can be made for a further, more costly and time-consuming evaluation.

References

Becht, M.C., Van Erp, C.F., Teeuwisse, T.M., Van Son, M., Van Heck, G.L., Van Son, M.J., Pop, V.J., 2001. Measuring depression in women around menopausal age: towards a validation of the Edinburgh Depression Scale. Journal of Affective Disorders 63, 209–213.

Bijl, R.V., Van Zessen, G., Ravelli, A., 1997. Psychiatric morbitidy among adults in The Netherlands: the NEMESIS-research II. Prevalence of psychiatric disorders (Psychiatrische morbiditeit onder volwassenen in Nederland: Het NEMESIS-onderzoek II. Prevalentie van psychiatrische stoornissen). Nederlands Tijdschrift voor Geneeskunde 141, 2453–2460.

Cox, J.L., Holden, J.M., Sagovsky, R., 1987. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. British Journal of Psychiatry 150, 782–786.

Cox, J.L., Chapman, G., Murray, D., Jones, P., 1996. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. Journal of Affective Disorders 39, 185–189.

Feldman, R., 2000. The ability of managed care to control health care costs: how much is enough? Journal of Health Care Finance 26, 15–25.

Glassman, H.D., Shapiro, P.A., 1998. Depression and the course of coronary artery disease. American Journal of Psychiatry 155, 4–11.

Harris, B., Huckle, P., Thomas, R., Johns, S., Funh, H., 1989. The use of rating scales to identify post-natal depression. British Journal of Psychiatry 154, 813–817.

Holder-Perkins, V.V., Wise, T., Williams, D.E., 2000. The somatizing patient. Current Psychiatry Reports 2, 234–240.

Kaufert, P.A., Gilbert, P., Tate, R., 1992. The Manitoba project: are-examination of the link between menopause and depression. Maturitas 14, 143–155.

Lave, J.R., Frank, R.G., Schulberg, H.C., Kamlet, M.S., 1998. Cost-effectiveness of treatments for major depression in primary care practice. Archives of General Psychiatry 55, 645–651.

Leonard, B.E., Miller, K., 1995. Stress and the Immune System: Immunological Aspects of Depressive Illness. Wiley, New York.

Murray, L., Carothers, A.D., 1990. The validation of the Edinburgh Post-natal Depression Scale on a community sample. British Journal of Psychiatry 157, 288–290.

Musselman, D.C., Evanc, D.C., Nemeroff, C.B., 1998. The relationship of depression and cardiovascular disease. Archives of General Psychiatry 55, 580–592.

Nyklíček, I., Louwman, W.J., Wijnands, C.J., Coebergh, J.-W.W., Pop, V.J., 2003. Depression and the lower risk for breast cancer development in middle-aged women: a prospective study. Psychological Medicine 33, 1111–1117.

Ormel, J., Schaufeli, W.B., 1991. Stability and change in psychological distress and their relationship with self-esteem and locus of control: a dynamic equilibrium model. Journal of Personality and Social Psychology 60, 288–299.

Panzarino Jr., P.J., 1998. The costs of depression: direct and indirect; treatment versus nontreatment. Journal of Clinical Psychiatry 20, 11–14.

Penninx, B.W., Guralnik, J.M., Pahor, M., Ferrucci, L., Cerhan, J.R., Wallace, R.B., Havlik, R.J., 1998. Chronically depressed mood and cancer risk in older persons. Journal of Clinical Endocrinology and Metabolism 90, 1888–1893.

Penninx, B.W., Geerlings, S.W., Deeg, D.J., van Eijk, J.T., van Tilburg, W., Beekman, A.T., 1999. Minor and major depression and the risk of death in older persons. Archives of General Psychiatry 56, 889–895.

Pop, V.J., Komproe, I.H., van Son, M.J., 1992. Characteristics of the Edinburgh Post natal Depression Scale in The Netherlands. Journal of Affective Disorders 26, 105–110.

Pop, V.J., Maartens, L.H., Leusink, G.L., 1998. Are autoimmune thyroid dysfunction and depression related? Journal of Clinical Endocrinology and Metabolism 83, 3194–3197.

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.

Schweiger, U., Deuschle, M., Korner, A., Lammers, C.H., Schmider, J., Gotthardt, U., Holsboer, F., Heuser, I., 1994. Low lumbar bone mineral density in patients with major depression. American Journal of Psychiatry 151, 1691–1693.

Smeets-Goevaers, C.G., Leusink, G.L., Papapoulos, S.E., 1998. The prevalence of low bone mineral density in Dutch perimenopausal women: The Eindhoven Perimenopausal Osteoporose Study. Osteoporosis International 8, 404–409.

Spitzer, R.L., Endicott, J., Robins, E., 1978. Research diagnostic criteria: rationale and reliability. Archives of General Psychiatry 35, 773–782.

Stoltz, C.M., Baime, M.J., Yaffe, K., 1999. Depression in the patient with rheumatologic disease. Rheumatic Disease Clinics of North America 25, 687–702.

WHO, 1998. WHO Report on Depression and Resource Utilization. Harvard University, Cambridge, MA.

Zich, J.M., Attkisson, C.C., Greenfield, T.K., 1990. Screening for depression in primary care clinics: the CES-D and the BDI. International Journal of Psychiatry in Medicine 20, 259–277.

Zonderman, A.B., Costa, P.T., McCrae, R.R., 1989. Depression as a risk for cancer morbidity and mortality in a nationally representative sample. Journal of the American Medical Association 262, 1191–1195.