Screening for depression in primary care

Development and validation of the Depression Scale, a screening instrument for depression


Depression is a common mental disorder; effective methods for treating it are also available. Its recognition and diagnosis are prerequisite to effective treatment. A majority of depressed patients are generally managed in the primary care setting; only a half of the cases, however, are identified at their first visit. Screening instruments to improve recognition of depression have therefore been developed. The Depression Scale (DEPS), consisting of 10 items, was developed and tested in primary care patients aged 18 to 64 years. Clinical assessments were made on the basis of Present State Examination interviews with 436 patients. The DEPS proved to be satisfactory. Increasing age and poor education had an adverse effect on the screening process, however. The sensitivity of the DEPS for clinical depression was 74%, and the specificity for non-depression 85%. The sensitivity for severe depression was 84%, and the specificity for symptom-free patients 93%. The DEPS seems to improve the recognition of depression in primary care and may also be suitable for screening depression in the general population and for identifying high-risk groups.

According to various studies, 10–19% of men and 21–34% of women suffer from depressive symptoms; the prevalence of clinical depression is 3–14% among men and 5–24% among women (1). In the last few years the prevalence of depression seems to have increased (2).

Depression is also common among patients consulting general practitioners. According to Barrett et al. (3), 10% of patients seen in general practice suffer from clinical depression and over 10% from milder depressive symptoms. Similar results have also been obtained by other researchers (4–8).

In Finland, over two thirds of the population visit a general community health center during a year (9), and psychiatric treatment is received by only 3–4% (10). According to Shepherd et al. (11), general practitioners refer only about 5% of patients to psychiatrists. A great majority of depressed patients are in fact treated in general practice; general practitioners thus need to be able to recognize and treat depression.

According to a number of studies, however, a considerable proportion of depression cases remain unidentified by general practitioners (12). It has been estimated that about one half of depressed patients are recognized by general practitioners at the first visit to the health care facility and another one tenth at subsequent visits, some 60% of cases being identified during the 6-month period following the first visit (8).

Approximately 70–80% of depressed patients show a favorable response to therapy (13). Recurrence of depression can also be prevented through therapeutic measures (14). According to recent studies, cases of depression identified by general practitioners have a better outcome than unrecognized depression (15–17); this finding, however, has not been confirmed in all studies (18).

Since a specific intervention in any disorder is not possible before its identification, the recognition of depression in primary care is a key to its effective treatment.

Use of a screening instrument as an aid in diagnosis

To facilitate the recognition of depression, various self-report screening scales (such as the Beck Depression Inventory (19, 20) and the Zung Self-rating...
Depression Scale (21) have been developed. Bech et al. (22, 23) have reviewed them and their applicability. Many of these scales are fairly lengthy, they have been developed primarily for psychiatric patient populations, and the instruments have often not been validated for the general population or for the patient population seen in general practice.

The Center for Epidemiologic Studies Depression Scale (CES-D) (24) is a screening device that has been used in many population studies. It consists of 20 questions, one half of which are positive and one half negative in character, with a total score ranging from 0 to 60 (24). In a study of the general population the CES-D had a sensitivity (correctly identified depressives) of 59% and a specificity (correctly classified non-depressives) of 86% (25). In particular, the false-negative rate was high (41%). Of the other instruments developed for screening depression in community populations, the Inventory to Diagnose Depression (IDD) (26) was developed for detection of major depressive disorders only.

Barrett et al. (27) developed a 20-item screening scale (with scores ranging from 0 to 60) suitable for use by general practitioners; it was derived from the Hopkins Symptom Checklist (HSCCL) (28) and the CES-D (24) and its questions concerning depressive symptoms are one-directional. When an abbreviated version of the Schedule for Affective Disorders and Schizophrenia (SADS) was used to detect depression in a sample of over 300 patients, with a cut-off score of 12, the screen had a sensitivity of 74% and a specificity of 87%.

Goldberg et al. (29) developed 9-item anxiety and depression scales for use by general practitioners and other non-psychiatrists. The full set of 9 questions needs to be administered only if there are positive answers to the first 4. According to the authors, patients with a depression score of 2 have a 50% chance of having a clinically important disorder.

The Depression Scale (DEPS)

The Tampere Depression Study (TADEP Study) was begun in 1990; it deals with the prevalence and clinical picture of depression in patients seen in community health centers and psychiatric treatment facilities. One objective of the study was to find out or develop a screening instrument suitable for the detection of depression in primary health care.

According to the literature, the screening instrument developed by Barrett et al. (27) was the only self-administered scale that was validated within patients seen in primary care and that had an acceptable correspondence with depression stated by a standardized interview method. It was therefore chosen as screening instrument for depression in this study.

The screen, supplemented with a question concerning insomnia, was tested in a pilot study. It was sent, together with a larger questionnaire, to a randomly selected sample of the general population aged 15 to 64 (a total of 653 individuals), to the patients of a community health center (199 individuals) and to the patients of a community mental health center (90 individuals). The questionnaire was returned by 428 individuals (273 of those in the community sample, 97 of patients receiving primary care and 58 of the patients of the mental health center).

The interview carried out to detect depression included the Present State Examination interview (9th version of the PSE) (30), the results of which were analyzed using the Catego program. All the interviewers had been trained in the use of the PSE. The subject was classified as depressive if he or she had an (Index of Definition) score of 5 or more and received an ICD diagnosis of 296 or 300.4.

This test of the screen showed that certain questions were poorly fulfilled or did not adequately differentiate PSE-diagnosed depression from other clinical disorders or from cases without diagnoses. Myers & Weissman (25) have also called attention to this issue. The authors decided therefore to shorten the screen by omitting items with a low response rate and with a poor ability to differentiate depression from other disorders. The differentiating capacity of the screen improved considerably on exclusion of questions nonspecific with regard to depression (for instance those concerning appetite, sleep disturbances, irritation, experiencing others as unfriendly, suicidal thoughts and sexual interest). This yielded a 10-item Depression Scale (DEPS) with a score ranging from 0 to 30; the questions of the Scale are shown in Table I. In the community health center sample of the pilot study, the DEPS had a sensitivity of 75% and a specificity of 88%.
Material and methods

The applicability and accuracy of the DEPS were investigated in the actual study phase of the TADEP project, using patients seen in community health centers. The study group consisted of 2486 individuals randomly selected from among individuals aged 18 to 64 who had visited one of three community health centers in Central Finland (Tampere, Kangasala and Lempäälä) or health care facilities operating under them (including consultations in normal office hours and out of hours, occupational health service and visits to prenatal clinics) between September 1991 and May 1992. During the visit, the patient was given the DEPS and another questionnaire and was asked to return them by mail. Of the patients who completed and returned the DEPS (n = 1643), all those who scored 9 or higher on the scale and 1 of 10 of those who were not picked up by the screen (i.e., with a score of <9) were interviewed. The interviews were conducted by 5 interviewers who had been trained in the use of the PSE and who had not participated in the selection of the interviewees. A total of 436 community health center patients were thus interviewed. The period between the arrival of the completed screening questionnaire and the interview was on average 15.3 days.

The TADEP project also included patients who sought care in the psychiatric treatment sector (community mental health centers, day hospital and psychiatric hospital) in the same geographical areas and during the same period (n = 563) and who completed the DEPS during the treatment visit. These results, however, are not dealt with here.

The definition of clinical disorder made use of the PSE (30), on which also the validation of the DEPS was based. In the interview, information was elicited on the occurrence of symptoms during both the preceding month and the preceding year. In the validation of the DEPS, only those experienced in the past month were considered.

Based on the PSE, 3 categories of depression were defined: (1) severe depression, consisting of cases with an ICD-8 diagnosis of 296, (2) mild depression, or cases with an ICD-8 diagnosis of 300.4 and with an Index of Definition of 5 or more, and (3) depressive symptoms, including PSE subtypes N+, O?N+, D?, M?, N?, O?R+ and R+ and an Index of Definition of 2–4. The other diagnoses and conditions with other symptoms were defined correspondingly.

Four of the 73 patients classified as having severe depression had a diagnosis of 296.1 (manic psychosis). All of them also had or had earlier had symptoms of depression; they were therefore classified under the heading of severe depression.

The internal validity of the DEPS was tested by calculating Pearson's correlation coefficients between the individual items and their sum scores and its internal reliability by calculating Cronbach's alpha for the whole scale. The external validity of the DEPS was analyzed by calculating the proportions of the patients suffering from the PSE depressions (severe or mild) or depressive symptoms in various DEPS score categories.

Results

Of the 1674 individuals who returned the questionnaire, 31 (2%) had not completed the screening scale. Those who had completed it had seldom left questions unanswered; 3% of the subjects had answered every question. The highest nonresponse rate for individual items was 4% (Table 1). Cronbach's alpha, which reflects the internal coherence of the screen, was 0.88. With the exception of the question concerning insomnia, the correlations of the item scores with the sum score of the screen were over \( r = 0.60 \) (Table 1).

There was no difference between the sexes in terms of completion of the screening questionnaire (number of items not responded to). Young subjects and those with upper secondary school education, on the other hand, had filled in the questionnaire more completely than others, i.e., they had fewer unanswered questions. ANOVA showed, however, that the association of education with completion of the questionnaire was explained by the subjects' age. When the effect of age (\( P = 0.009 \)) was taken into account, basic education was not significantly related to completion of the scale. There was likewise no interaction between age and education.

The discriminating ability of the DEPS was examined by comparing the scores obtained for the various samples in the pilot study and the actual TADEP project (Table 2). The screen score for the community sample was lower than that for the patients of community health centers (\( P = 0.05 \)), who in turn had lower scores (\( P < 0.001 \)) than psychiatric

<table>
<thead>
<tr>
<th>Sample</th>
<th>General population mean</th>
<th>SD</th>
<th>Primary care mean</th>
<th>SD</th>
<th>Psychiatric outpatient care mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piilot study</td>
<td>n=273</td>
<td>4.99 (5.14)</td>
<td>5.84 (5.41)</td>
<td>13.43 (7.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(15–64 years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TADEP study</td>
<td>n=1643</td>
<td>5.59 (5.02)</td>
<td>12.11 (6.96)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(18–64 years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
outpatients. The differences were quite clear and in keeping with expectations.

The condition clearly best identified by the DEPS was severe depression. Only about 14% of these cases remained unrecognized, while of cases of mild depression, as many as 37% remained unidentified (Table 3). Of the patients with some depressive symptoms, slightly under one half were picked up by the screen, while only about one tenth of subjects with other symptoms were picked up. The number of completely asymptomatic persons picked up by the screen was low; the specificity of the screen in the case of this group was as high as 93%. At the 9-point cut-off, the DEPS had a sensitivity, i.e., the percentage of correctly diagnosed cases of clinical depression, of 74% and a specificity, i.e., the percentage of correctly identified cases of non-depression, of 85%.

In terms of sensitivity and specificity, the screen functioned equally well with men and women. With advancing age, the screen worked somewhat less well. The effect of education was greater, however: the sensitivity for subjects with lower secondary education was 69% and the specificity 83%, while the figures for those with upper secondary education were 100% and 90%, respectively.

The proportions of subjects assessed in the PSE interview as suffering from depressive illness or depressive symptoms in various DEPS score categories are shown in Fig. 1.

After the DEPS score exceeded 9, the combined percentages of diagnoses of depression rose sharply; the proportion of cases of mild depression, however, was highest in the group with the highest scores. The relative proportion of individuals with depressive symptoms was highest in the group with scores in the range 6–11.

In Fig. 1 attention is drawn to one subject who scored 28 on the DEPS but did not meet the criteria for clinical depression in the PSE interview. Some month before the study, the subject had lost his job, causing him considerable financial problems. At time the administration of the DEPS the subject suffered from heavy anxiety and depressive symptoms. Soon after screening he got a new job and by the time of the interview the subject felt well. In the PSE inter-

![Fig. 1. Proportions of patients with severe or mild depression or depressive symptoms according to DEPS scores.](image-url)
view the subject reported having experienced practically no symptoms in the past month; in the part of the interview concerning symptoms experienced during the preceding year, however, he was diagnosed as having severe depression (296).

Discussion

An instrument suitable for screening of depression has to satisfy many conditions. Since cognitive disorders occur in association with depression (31, 32), the screening questionnaire should as brief as possible, should be easy to complete in a short time and should consist of questions which are clear and unambiguous. The questions should be compatible with the respondent's experiential world and culturally acceptable from his or her point of view. Finally, a depression-screening instrument must naturally be able to differentiate as accurately as possible between depression and non-depression.

The pilot study showed that the only available self-administered screening instrument that had been validated within patients seen in primary care did not function adequately enough within the Finnish primary care population. For instance, questions concerning self-destructiveness, sexuality and experiencing others as unfriendly were answered by Finnish subjects less frequently. On the other hand, there were questions that did not differentiate depression adequately from other disorders or from patients with no diagnosis. These findings gave us a basis for working up the instrument for more suitable in this patient group.

The final study showed that the DEPS fulfilled the requirements of screening instrument quite satisfactorily. It is one of the shortest self-administered screening instruments for depression, consisting of 10 questions only. The number of questions left unanswered in the case of each item remained low, and the completion rate was not affected by the subject's sex. With advancing age, however, the number of unanswered questions rose, although the subjects were under 65. The reason for this may lie in an age-related impairment of cognitive functions (33) and its reflection in the form of some difficulty in completing the questionnaire. It is also possible that depressed mood, well known to involve cognitive disorders (31, 32), is reflected in the form of difficulty in filling in the screening questionnaire. The fact that the mean DEPS scores — similarly to the number of uncompleted DEPS items — rose with advancing age points in this direction. This question has not previously been given adequate consideration. It seems evident, however, that older patients, in particular, may need guidance in filling in the questionnaire.

The subjects' level of education also seems to have a certain effect on the accuracy of the screen, its differentiating capacity. A high level of basic education was related to a high specificity and sensitivity of the scale. Better education evidently makes it easier to understand the questions of the scale and thereby increases its clinical relevance. The effect of education should be taken into account in screening for depression.

Despite the above limitations, which have received little attention in screening studies of depression, the DEPS seemed to pick up cases of clinical depression quite satisfactorily. Its sensitivity (74%) and specificity (85%) were of the same order of magnitude as the figures attained by Barrett et al. (27) (74% and 87%, respectively). The sensitivity of the Goldberg et al. (29) depression scale in severe depressive episode was 85% and the specificity in patients with no disorder 91%. In the latter study the inquiry concerning the symptoms included in the scale and the subject's diagnostic assessment was, however, carried out by the same psychiatrist. In the present study, in which the PSE interviewers did not participate in the screening process or in the selection of interviewees, the sensitivity in severe depression was 86% and specificity in subjects with no diagnosis or symptoms 93%. The DEPS thus seems to work at least as well as the above scales.

It has been estimated that about one half of cases of depression are identified by general practitioners at the first treatment visit (8). The DEPS was able to identify about three fourths of all cases of clinical depression and as many as 86% of cases of severe depression. It may thus be concluded that the DEPS can fundamentally improve recognition of depression in patients seen in community health centers.

Only a small number (7%) of the completely asymptomatic patients scored 9 or higher. In addition to depressive illness and depressive symptoms, the screen picked up individuals with other symptoms and diagnoses. The screening should thus in positive cases be followed by a careful diagnostic interview by the attending physician and a thorough exploration of the patient's life situation. A false-positive screening diagnosis could thereby benefit the patient rather than cause any harm. A certain number of such patients suffer from rapid, transient fluctuations of mood (34). Their follow-up, even without immediate intervention, may be useful. A considerable proportion of false-positives, on the other hand, suffer from other mental disorders or symptoms, above all anxiety, and a closer exploration of the problems is indicated before specific therapy is initiated.

Thus the DEPS can be a useful aid for the general practitioner in clinical work, but it cannot be used as the only diagnostic criterion and the sole
basis for initiation of therapy. It can at best draw the physician’s or other care provider’s attention to the possibility of depression, lead to a more thorough examination of the patient and thereby improve diagnostic accuracy. A carefully taken history and a thorough examination of the mental status of the patient is required to arrive at a correct diagnosis (8).

The examination of the depressed patient in fact requires not only skill but often also time and an ability to listen to the patient on his own terms (34, 35). In clinical practice, though, treatment is often provided without a clear diagnosis of depression (36).

In addition to screening for depression, the DEPS may also be useful in the following of the prevalence of depression and its changes in the population (37) and in the identification of groups at high risk for depression. Only after more experience is gained with the DEPS, however, can its usefulness in these applications be determined.

Acknowledgements

We thank doctors and other personnel of the Health Centres in Tampere, Kangasala and Lemmäla for the excellent cooperation in data collection. The Academy of Finland supported the study financially.

References


Screening for depression in primary care


