The Depression Scale as a screening instrument for a subsequent depressive episode in primary healthcare patients

OUTI POUTANEN, ANNA-MAIJA KOIVISTO, MATTI JOUKAMA, AINO MATTILA and RAIMO K. R. SALOKANGAS

Background There are numerous instruments for screening for depression. A feasible screen is good at both recognising and predicting depression.

Aims To study the ability of the Depression Scale and its items to recognise and predict a depressive episode.

Method A sample of patients attending primary care was examined in 1991–1992 and again 7 years later. The accuracy of the Depression Scale at baseline and at follow-up was tested against the Short Form of the Composite International Diagnostic Interview (CIDI–SF) diagnosis of depression at follow-up. The sensitivity and specificity of the Depression Scale and its items were assessed.

Results Both baseline and follow-up Depression Scale scores were consistent with the CIDI–SF diagnoses. It was possible to find single items efficient at both recognising and predicting depression.

Conclusions The Depression Scale is a useful screening instrument for depression, with both diagnostic and predictive validity.

Declaration of interest None. Funding from the Medical Research Fund of Tampere University Hospital.

There are several instruments to help primary care clinicians identify patients with major depression (Williams et al, 2002). The Depression Scale (Salokangas et al, 1995) is one of these. The relatively low prevalence of depression in primary care practice requires that the sensitivity and specificity of a screening instrument should be almost perfect (Schwenk, 1996). The Beck Depression Inventory (BDI; Beck et al, 1961) and the Hospital Anxiety and Depression Scale (HADS; Zigmond & Smith, 1983) are the most commonly used screening instruments. The popularity of a scale does not guarantee that it is feasible and up-to-date (Bagby et al, 2004). In this study, we aimed to examine the ability of the Depression Scale and its items to recognise and predict a depressive episode.

METHOD

This study forms a part of the larger Tampere Depression Project (TADEP), the baseline study of which was done in 1991–1992 (Salokangas et al, 1995, 1996; Salokangas & Poutanen, 1998). Consecutive patients aged 18–64 years attending primary care services (including consultations in normal office hours and out of hours, occupational health services and visits to prenatal clinics) completed a postal questionnaire including questions on their demographic characteristics, health and functioning, as well as a screening instrument for depression (the Depression Scale; Salokangas et al, 1995). Of the 1643 patients who returned the screening questionnaire adequately filled in, all who screened positive for depression (n=372) and every tenth person who was screen-negative (127 out of 1271 individuals) were invited for interview. To diagnose clinical depression, the Present State Examination (PSE; Wing et al, 1974) was used. A total of 436 persons were interviewed. Their PSE diagnoses were as follows: severe depression n=55, depressive symptoms n=60, other psychiatric symptoms n=174, other psychiatric diagnosis n=29, no psychiatric symptom n=55.

Seven years later a follow-up study was conducted. The number of participants to whom the follow-up questionnaire could be posted was 413 (11 people were dead, no address could be found for 6 and 6 others had attended psychiatric out-patient care and were excluded from subsequent analysis in the present primary care study). Of these 299 returned the questionnaire, and 250 (57.3% of the baseline sample) were willing to take part in the telephone interview. Men (P=0.050) and married individuals (P=0.018) participated more frequently than women or those who were not married. The study protocol was approved by the Tampere University Hospital ethics committee and written informed consent was obtained from the participants.

Study procedure

The Depression Scale includes ten items, with four response alternatives scoring 0–3: ‘not at all’, ‘a little’, ‘quite a lot’ and ‘extremely’ (see Table 2). In the baseline study the cut-off point for the screening sum score was >8.

In the follow-up study participants again filled in the Depression Scale, the Michigan Alcoholism Screening Test (Selzer, 1971), parts of the Hopkins Symptom Checklist (Derogatis et al, 1974), and structured questions. To assess major depressive episode, 38 items from the Short Form of the Composite International Diagnostic Interview (CIDI–SF; World Health Organization, 1989; Kessler et al, 1998) were used in a telephone interview. The CIDI–SF questions concerning the occurrence of symptoms of a major depressive episode referred to the previous month. Three trained psychiatrists (A.M. and Drs Liisa Groth and Niko Seppälä), each with at least 5 years’ experience in psychiatry, conducted the interviews, masked to the baseline PSE diagnoses.

Statistical methods

The accuracy of the Depression Scale as a screening instrument for depression was assessed by receiver operating characteristic (ROC) curve analyses. The follow-up Depression Scale score (DEPS–F) was compared with the CIDI–SF diagnosis of depression. The ability of the baseline Depression Scale score (DEPS–B) to predict
the CIDI–SF diagnosis at follow-up was also evaluated. In ROC analyses, sensitivity, specificity and areas under the curve were calculated. Sensitivity and specificity were calculated for each reasonable cut-off point of the Depression Scale.

To evaluate which single items of the DEPS–B were best at predicting a depressive episode, the sensitivity and specificity for single items were calculated. After that, logistic regression analysis with forward stepwise method using all DEPS–B items as predictors was conducted. For this analysis, all items were dichotomised using 1 as the cut-off score (0–1, negative item result; 2–3, positive item result). To evaluate which single items of the DEPS–F were best for recognising a depressive episode, the sensitivity and specificity were calculated separately for each item, and logistic regression analysis was likewise conducted.

To identify an ideal pair of Depression Scale items for composing a short version of both DEPS–B and DEPS–F, sensitivity and specificity for every possible DEPS–B and DEPS–F item pair were calculated. An ideal pair of items implied that both of the items scored above 1. Only pairs in which sensitivity was at least 50% were regarded as relevant and reported.

Analyses were performed using the Statistical Package for the Social Sciences version 11.5 for Windows; P<0.05 was considered statistically significant.

**RESULTS**

**Depression Scale v. CIDI–SF**

In participants with CIDI–SF depression, the median DEPS–F score was 18 (range 7–30) and in those without depression it was 5 (range 0–28) (P<0.001, Mann–Whitney test). In the ROC analysis of DEPS–F v. CIDI–SF the area under the curve was 0.939 (Fig. 1). The ideal pair of sensitivity (90.5%, 95% CI 0.71–0.97) and specificity (86.8%, 95% CI 0.82–0.91) was found with a score of >11 as the cut-off point (Table 1). In participants with CIDI–SF depression the median DEPS–B score was 17 (range 2–24) and in those without depression it was 10 (range 0–28) (P<0.001, Mann–Whitney test). In the ROC analysis of DEPS–B v. CIDI–SF the area under the curve was 0.803 (Fig. 1). The ideal pair of sensitivity (72.7%,

### Table 1  Sensitivity and specificity of different Depression Scale cut-off points

<table>
<thead>
<tr>
<th>Depression scale score</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score at follow-up v. CIDI–SF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>95.2</td>
<td>74.4</td>
</tr>
<tr>
<td>9</td>
<td>95.2</td>
<td>78.5</td>
</tr>
<tr>
<td>10</td>
<td>90.5</td>
<td>83.6</td>
</tr>
<tr>
<td>11</td>
<td>90.5</td>
<td>86.8</td>
</tr>
<tr>
<td>12</td>
<td>81.0</td>
<td>89.5</td>
</tr>
<tr>
<td>13</td>
<td>76.2</td>
<td>91.3</td>
</tr>
<tr>
<td>14</td>
<td>71.4</td>
<td>92.2</td>
</tr>
<tr>
<td>15</td>
<td>71.4</td>
<td>94.1</td>
</tr>
<tr>
<td>Score at baseline v. CIDI–SF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>95.5</td>
<td>27.3</td>
</tr>
<tr>
<td>9</td>
<td>90.9</td>
<td>41.2</td>
</tr>
<tr>
<td>10</td>
<td>86.4</td>
<td>53.7</td>
</tr>
<tr>
<td>11</td>
<td>86.4</td>
<td>62.5</td>
</tr>
<tr>
<td>12</td>
<td>72.7</td>
<td>71.8</td>
</tr>
<tr>
<td>13</td>
<td>72.7</td>
<td>77.8</td>
</tr>
<tr>
<td>14</td>
<td>63.6</td>
<td>83.8</td>
</tr>
<tr>
<td>15</td>
<td>59.1</td>
<td>87.5</td>
</tr>
</tbody>
</table>

CIDI–SF, Composite International Diagnostic Interview–Short Form; DEPS, Depression Scale.

### Table 2  Sensitivity and specificity of Depression Scale items at baseline and at follow-up compared with depression assessment with the Composite International Diagnostic Interview.

<table>
<thead>
<tr>
<th>Depression Scale items</th>
<th>DEPS score at follow-up</th>
<th>DEPS score at baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>During the past month I have . . .</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. . . suffered from insomnia</td>
<td>63.6</td>
<td>84.4</td>
</tr>
<tr>
<td>2. . . felt blue</td>
<td>59.1</td>
<td>89.3</td>
</tr>
<tr>
<td>3. . . felt everything was an effort</td>
<td>81.8</td>
<td>86.4</td>
</tr>
<tr>
<td>4. . . felt low energy or slowed down</td>
<td>72.7</td>
<td>83.8</td>
</tr>
<tr>
<td>5 . . felt lonely</td>
<td>59.1</td>
<td>93.9</td>
</tr>
<tr>
<td>6. . . felt hopeless about the future</td>
<td>81.8</td>
<td>92.5</td>
</tr>
<tr>
<td>7 . . not got any fun out of life</td>
<td>54.5</td>
<td>84.8</td>
</tr>
<tr>
<td>8 . . had feelings of worthlessness</td>
<td>50.0</td>
<td>96.1</td>
</tr>
<tr>
<td>9 . . felt all pleasure and joy has gone from life</td>
<td>45.5</td>
<td>93.8</td>
</tr>
<tr>
<td>10 . . felt that I cannot shake off the blues even with help from family and friends</td>
<td>42.9</td>
<td>91.2</td>
</tr>
</tbody>
</table>

CIDI–SF, Composite International Diagnostic Interview–Short Form; DEPS, Depression Scale.

1. All items are scored 0, not at all; 1, a little; 2, quite a lot; 3, extremely. An item was regarded as positive when the score was >1.
Depression Scale items at baseline and at follow-up from logistic regression analyses significantly associated with depression at follow-up assessment.

<table>
<thead>
<tr>
<th>Depression Scale items</th>
<th>OR</th>
<th>(95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPS at follow-up v. depression at follow-up (CIDI–SF)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During the past month I have . . .</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 . . . felt everything was an effort</td>
<td>5.54 (1.35–22.79)</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>6 . . . felt hopeless about the future</td>
<td>21.89 (5.45–88.01)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>DEPS at baseline v. depression at follow-up (CIDI–SF)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During the past month I have . . .</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 . . . suffered from insomnia</td>
<td>2.67 (0.99–7.19)</td>
<td>0.055</td>
<td></td>
</tr>
<tr>
<td>3 . . . felt everything was an effort</td>
<td>6.50 (1.76–24.01)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>9 . . . felt all pleasure and joy has gone from life</td>
<td>3.70 (1.35–10.09)</td>
<td>0.011</td>
<td></td>
</tr>
</tbody>
</table>

CIDI–SF, Composite International Diagnostic Interview–Short Form; DEPS, Depression Scale.

Table 4  Sensitivity and specificity of Depression Scale item pairs at baseline and at follow-up compared with depression at follow-up assessment

<table>
<thead>
<tr>
<th>DEPS item pair¹</th>
<th>DEPS score v. CIDI–SF episode of depression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DEPS at follow-up</td>
</tr>
<tr>
<td></td>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>1+3</td>
<td>59.1</td>
</tr>
<tr>
<td>1+6</td>
<td>54.5</td>
</tr>
<tr>
<td>2+3</td>
<td>50.0</td>
</tr>
<tr>
<td>2+6</td>
<td>59.1</td>
</tr>
<tr>
<td>2+9</td>
<td>27.3</td>
</tr>
<tr>
<td>3+4</td>
<td>72.7</td>
</tr>
<tr>
<td>3+5</td>
<td>54.5</td>
</tr>
<tr>
<td>3+6</td>
<td>72.7</td>
</tr>
<tr>
<td>3+9</td>
<td>45.5</td>
</tr>
<tr>
<td>4+6</td>
<td>68.2</td>
</tr>
<tr>
<td>5+6</td>
<td>54.5</td>
</tr>
</tbody>
</table>

CIDI–SF, Composite International Diagnostic Interview–Short Form; DEPS, Depression Scale.

The first validation of the Depression Scale was reported in an earlier study, in which the cut-off point for depression was > 8 (Salokangas et al., 1995). In the baseline validation study, using the PSE as the criterion, the sensitivity of the Depression Scale for clinical depression was 74% and the specificity for non-depression 85%. For severe depression the figures were 84% and 93%. In the present study the figures for sensitivity and specificity were better than those of the earlier validation study. In the baseline validating analyses the sampling ratio was taken into account, but this was not done in the present study, which was only intended to ascertain the ability of the scale to predict an episode of depression and to evaluate its individual items. The differences in the levels of sensitivity and specificity between the baseline validation analyses and these follow-up analyses are perhaps partly explained by this fact. There are also differences in the validity criterion between the two diagnostic instruments. The PSE is based on symptoms, and the CIDI is based on syndromes (Lowe et al., 2004). With the CIDI–SF the definition of depression was clearer because there were only two categories: depressive and non-depressive. It should also be kept in mind that the PSE interviews at baseline were held face-to-face, whereas the CIDI–SF interviews at follow-up were the best balance of recognition and prediction. Only the pairs with sensitivity of at least 50% are reported (Table 4). The three best pairs for recognition were items 3 and 6, items 3 and 4, and items 4 and 6, whereas the best pairs for prediction were items 2 (‘I have felt blue’) and 3, items 3 and 4, and items 3 and 9.

Discussion

The Depression Scale was quite consistent with the CIDI–SF both as a predictor and a recogniser of depression. ‘Feeling that everything is an effort’ and ‘feeling hopeless about the future’ were the best items, and also the best item pair for recognising depression. ‘Suffering from insomnia’ ‘feeling everything is an effort’ and ‘feeling all pleasure and joy were gone from life’ were the best items for predicting future depression. ‘Feeling blue’ and ‘feeling everything is an effort’ were the best item pair for predicting future depression.

Sensitivity and specificity

The three most sensitive DEPS–F items were 3 (‘I have felt everything was an effort’), 6 (‘I have felt hopeless about the future’) and 4 (‘I have felt low energy or slowed down’), and the most specific items were 8 (‘I have had feelings of worthlessness’), 5 (‘I have felt lonely’) and 9 (‘I have felt all pleasure and joy has gone from life’) (Table 2). In the case of DEPS–B, item 3 had a high sensitivity whereas items 9, 5, 8 and 10 (‘I felt that I cannot shake off the blues even with help from family and friends’) had a reasonably high specificity. One item (item 3) was quite sensitive in both analyses, for both recognising and predicting CIDI–SF depression.

In logistic regression analyses, DEPS–F items 3 and 6 were significantly associated with CIDI–SF depression, whereas DEPS–B items 1 (‘I have suffered from insomnia’), 3 and 9 significantly predicted occurrence of subsequent CIDI–SF depression (Table 3).

Best Depression Scale item pairs v. CIDI–SF

Sensitivity and specificity were calculated for every possible pair of Depression Scale items to ascertain which two items had

95% CI 0.52–0.87) and specificity (77.8%, 95% CI 0.72–0.83) was found with a score of >13 as the cut-off point (Table 1).

Depression Scale items v. CIDI–SF

The three most sensitive DEPS–F items were 3 (‘I have felt everything was an effort’), 6 (‘I have felt hopeless about the future’) and 4 (‘I have felt low energy or slowed down’), and the most specific items were 8 (‘I have had feelings of worthlessness’), 5 (‘I have felt lonely’) and 9 (‘I have felt all pleasure and joy has gone from life’) (Table 2). In the case of DEPS–B, item 3 had a high sensitivity whereas items 9, 5, 8 and 10 (‘I felt that I cannot shake off the blues even with help from family and friends’) had a reasonably high specificity. One item (item 3) was quite sensitive in both analyses, for both recognising and predicting CIDI–SF depression.

In logistic regression analyses, DEPS–F items 3 and 6 were significantly associated with CIDI–SF depression, whereas DEPS–B items 1 (‘I have suffered from insomnia’), 3 and 9 significantly predicted occurrence of subsequent CIDI–SF depression (Table 3).

Best Depression Scale item pairs v. CIDI–SF

Sensitivity and specificity were calculated for every possible pair of Depression Scale items to ascertain which two items had
conducted by telephone. A telephone inter-
view relies more on the examinee’s own
assessment, and is closer to a self-rating
instrument like the Depression Scale. The
same items of the CIDI–SF were used as
in a previous Finnish depression study
(Isometsä et al, 1997; Lindeman et al,
2000) using the computer-assisted telephone
interview method.

According to Lowe et al (2004) the
sensitivity of screening questionnaires
should lie above specificity and be as high
as possible, and the specificity should be
at least 75%. In this study the cut-off point
>11, which has a sensitivity of 90.5% and
specificity of 86.8%, could be ideal.

When the ability of the Depression
Scale to predict an episode of depression
was analysed, the area under the curve
was 0.803. An earlier study with primary
care patients (Salokangas et al, 1994)
showed that the rate of clinical depression
in people with a Depression Scale score
above 12 was about 47% and in those with
a score above 15 it was about 57%. These
percentages are high enough to have some
clinical value. In this study, with a cut-off
point of >11 sensitivity was 86.4% but
specificity only 62.5%. When an instru-
ment is used as a predictor it is perhaps
more important to avoid false positives
and not to stigmatise patients; this justifies
a higher cut-off point.

What did the Depression Scale
actually assess?

In a study of general practice patients
(Williamson et al, 2005), four mental
health self-report scales and a composite
of those four were assessed to determine
their accuracy in predicting psychiatric
caseness for depression, dysthymia, gener-
alised anxiety disorder, social phobia, agor-
aphobia and panic attack. One scale
measuring neuroticism – the Neuroticism
Scale of the Eysenck Personality Question-
naire (EPQ–N; Eysenck et al, 1985) – and
a composite of all four scales were found
to be very strong and accurate predictors
of psychiatric caseness, but they were
unable to differentiate between specific
disorders. In our study only episode of
depression – not other psychiatric diag-
noses – was assessed.

In an extensive follow-up study (Tyrer
et al, 2004) the quick-to-use HADS was
good for recognising both depression and
anxiety, and was better than any other
single measure for predicting the outcome
of both anxiety and depressive disorders
after an interval of 12 years. The Mont-
gomery–Åsberg Depression Rating Scale
did not have such predictability.

When the Depression Scale and two
common self-rating instruments (the BDI
and the HADS) are compared, they differ
in many ways. The Depression Scale con-
centrates on the previous month, whereas
the BDI concentrates on the previous week
(the BDI–II on the past 2 weeks; Beck et al,
1996) and the HADS on current feelings.
Of the criterion standards used in this
study, both the PSE and the CIDI–SF refer
to the previous month. It is difficult to
say, however, what the true significance
of the differences in these time periods is.

The Depression Scale is the shortest of
the three instruments, and the BDI is the
longest. The formulation of the items is
different: the most evident difference is
that the Depression Scale gives exactly the
same short-answer alternatives for all ten items,
whereas there are several different sets of
alternative answers in both the BDI and
the HADS. This makes the Depression
Scale very quick and easy to use, and
increases adherence.

The BDI includes most of the Depres-

sion Scale topics. Only the topics of items
5 (loneliness), 7 (no fun) and 10 (not helped
even with family and friends) are missing in
the BDI. The Depression Scale item 5 was
specific in recognising depression and item
10 specific in predicting it. However, the
BDI covers the symptoms of depression
more comprehensively than the former
scale. The HADS covers both depression
and anxiety, but lacks most of the Depres-
sion Scale topics (items 1, 2, 3, 5, 8 and 10);
the symptoms covered are less severe
than in the BDI or in the Depression Scale.
Common topics for all the three self-rating
instruments are the Depression Scale items
4 (low energy), 6 (hopelessness), and 9 (lost
pleasure and joy). These topics probably
relate to the core of depression symptom-
tology; other topics can be said to be conse-
quences of the core symptoms and not so
essential to depression only.

The Depression Scale items 3 and 4
were good at both recognising and predict-
ing depression. Item 3 (‘I have felt every-
thing was an effort’) suggests reduction of
energy, which is one of the main symptoms
of depression according to the ICD–10.
Item 6 was good for recognition even
though its wording refers to the future (‘I
have felt hopeless about the future’); hope-
lessness is also a symptom of depression in
the ICD–10. Item 9 was good in predicting
depression. The wording of item 9 (‘I have
felt all pleasure and joy has gone from life’)
refers to something that has already hap-
pened, something that is possibly endured
as beyond help. Item pair 2 and 3 was the
best at predicting depression. The wording
of item 2 (‘I have felt blue’) may be experi-
enced as persistent low mood, referring to
a more chronic state. It is almost the same as
lowering of mood, one of the main symp-
toms of depression in ICD–10. The best
combination – and a possible quick version
– of two items for recognising depression
was items 3 and 6, and the best com-
bination for predicting depression was
items 2 and 3.

The use of psychometric scales is in gen-
eral problematic. Among people who ap-
pear to be healthy according to standard
mental health scales it is possible to identify
a subgroup of people who may not be
psychologically healthy at all: mental
health scales may assess not mental health
but instead defensive denial (Shedler et al,
1993). Moreover, any scale that is valid
for assessing current depression will have
some long-term predictability because de-
pression is recurrent. However, if a scale
has predictability, it means it has the ability
to catch not just reactive and short-term
symptoms but more chronic or recurrent
core features of the disorder.

Limitations and strengths
of the study

It is a limitation of the study that the inter-
views were held by telephone. However,
the CIDI–SF telephone interviews were
conducted with care and by experienced
psychiatrists. Some information about the
mental state of these patients during the fol-
low-up period was gathered, but this was
self-report information and possibly not
so reliable, and we decided not to use it in
this study. This was not a follow-up study
in its truest sense: the assessments were
made only twice – at baseline and 7 years
later. Thus, the mental state of the par-
ticipants during the intervening period
is obscure, decreasing slightly the credibility
of the study. It is strength of the study that
the sample was fairly large, and that it was
a follow-up study with a wide range of
primary care patients.

Implications

The Depression Scale is not only an easy-to-
use screening instrument, it also appears to
be a reasonably good predictor for a depressive episode years ahead. It seems to work well with patients who have vague psychiatric symptoms, as is often the case in primary healthcare. Some of its items have a better ability to recognise or to predict depression than others; this suggests the possibility of creating an even shorter version of this scale.

REFERENCES


