

The origins and development of the Edinburgh Postnatal Depression Scale

The development of the Edinburgh Postnatal Depression Scale (EPDS) was first described by Cox *et al* in 1987 and was subsequently summarised in *Perinatal Psychiatry* (Cox & Holden, 1994). In these publications we suggested that the EPDS would help in identifying postnatal depressive disorders which would otherwise be undetected in the community. Prospective studies of postnatal depression in Uganda and Scotland had found frequencies of 10% and 13% respectively (Cox, 1983). These were striking findings which at the time increased awareness of the need to develop methods to detect this disabling mood disorder. Depression in the African mothers living in villages north of Kampala prevented them from working (digging and carrying water) and most received no help at all from traditional healers or health professionals. The Scottish women also struggled to meet the demands of their baby and their partner.

Existing scales

It was apparent that existing self-report scales for depression were unlikely to be useful in detecting depression in childbearing women. The State of Anxiety and Depression (SAD) self-report scale of Bedford & Foulds (1978), the Beck Depression Inventory (BDI; Beck *et al*, 1961) and the General Health Questionnaire (GHQ; Goldberg, 1972) all had serious limitations for use with pregnant and post-partum women. Women might endorse (tick) the somatic items on the scales because of the physiological changes of childbearing (e.g. weight gain, breathlessness and tachycardia), and childbearing women can disclose normal worries. Sleep difficulty as a symptom of depression is difficult to evaluate when sleep is being disturbed by the baby.

In 1983 we noted that such 'false positives' in self-report questionnaires might reduce the reliable detection of neurosis in pregnant and post-partum women, and that scales specifically for use during pregnancy and in the puerperium might be needed (Cox, 1983). Snaith (1983), who had developed the Hospital Anxiety and Depression (HAD) scale (Zigmond & Snaith, 1983), also recognised the need to modify existing self-report scales for use in specific clinical situations. Williams *et al*

(1980) had emphasised that questionnaires validated for use on hospital samples should be revalidated when used in the community. Thus, by the mid-1980s the need to develop a depression scale specifically validated for use by childbearing women was apparent and increasingly compelling.

Validating the Edinburgh Postnatal Depression Scale

At the outset we recognised that a questionnaire for use with childbearing women would need to be simple to complete and acceptable to people who did not regard themselves as unwell. Furthermore, the healthcare worker administering the scale might not have had any specialised training in psychiatric disorders. Finally, the new scale would need to have satisfactory validity and reliability, and be sensitive to changes in the severity of depression over time.

Clinical experience when assessing and treating women with postnatal depression was used to identify possible items from questionnaires such as the SAD and HAD scales and the BDI. Items that lacked 'face validity' (i.e. that would not have been understood by childbearing women or might have been inappropriate at this time, e.g. 'I can enjoy a good book or radio or television' or 'I feel as if I am slowed down') and 'somatic' items (misleading as indicators of depression) were discarded. We then selected 30 items, which included several of our own construction, to pilot with women who were asked to comment about the wording and the order of the items.

We eventually agreed on 13 items that we thought likely to detect mothers with clinical depression. The resultant questionnaire was then validated (Cox, 1986) on a sample of 60 postnatal women, who completed the 13-item scale and were interviewed by the psychiatrist Ruth Sagovsky using Goldberg *et al's* (1970) Clinical Interview Schedule. The diagnosis of depression using major and minor Research Diagnostic Criteria (RDC; Spitzer *et al*, 1978) was then established. The 13-item scale was found to distinguish satisfactorily between women with and without depression. A factor analysis, however, showed not only a 'depression' factor that explained 46% of the variance but also another factor that loaded on three items ('I have enjoyed being a mother' and the two irritability items). We therefore realised that the 13-item scale could be shortened to 10 without impairing its effectiveness. This shortened 10-item scale (named the Edinburgh Postnatal Depression Scale) had no specific item about mothering the baby nor about irritability, a development that later widened the potential use of the scale to other populations.

A second validation of the 10-item version was carried out on a sample of 84 postnatal women who were taking part in an ongoing

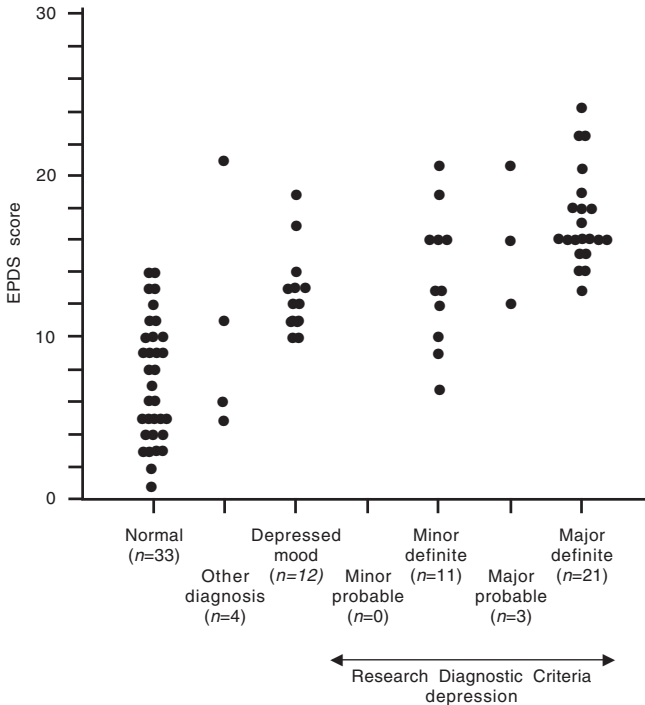


Fig. 2.1 Validation of the 10-item Edinburgh Postnatal Depression Scale (EPDS; Cox *et al*, 1987).

study of health visitor counselling (Fig. 2.1) (Cox *et al*, 1987). The shortened scale identified all women with a definite major depression and two of the three with a probable major depression at a cut-off of 12/13. Of the 11 women with a definite minor depression, only four had a false-negative score. Although this cut-off resulted in 11 false positives, 6 of these women had several depressive symptoms without fulfilling all RDC for clinical depression. Interestingly, three women with a psychiatric diagnosis other than depression all scored below the cut-off.

The sensitivity of the EPDS (the proportion of women with RDC depression who were true positives) was 86%, and the specificity (proportion of RDC non-depressed women who were true negatives) was 78%. The positive predictive value (the proportion of women above the threshold on the EPDS ($n = 41$) who met RDC for depression ($n = 30$)) was 73%. These findings suggested that the rate for failing to detect women with depression could be reduced to under 10% by using a lower cut-off, of 9/10. This was the cut-off that we recommended in our initial publication of the EPDS for use as a first-stage screening measure.

Reliability

The split-half reliability of the 10-item EPDS was 0.88 and the standardised α coefficient 0.87. (The split-half reliability was obtained by dividing items on the scale into two halves, which were then compared; a high correlation suggested that the items were measuring the same characteristics.) The sensitivity to change in severity of depression over time was established by comparing EPDS scores obtained at two interviews separated by a 3-month interval. This showed that the EPDS can be used to detect changes in the level of depression over time, so widening its use as an outcome measure in intervention studies.

The three women with a false-negative score each had other family members present when they were interviewed. We recommend that the EPDS is optimally completed when other family members are not present, as women may exaggerate or minimise their problems in these circumstances. The 10-item EPDS was found to be very acceptable to most of the women and, as important, to their health visitors. Another advantage of the scale was its brevity and simplicity of scoring.

These findings, together with our clinical experience, suggested that the EPDS would be useful in detecting postnatal depression in the community as well as in research projects. A cut-off of 9/10 was likely to detect almost all cases of depression, with very few false negatives. This cut-off score is particularly useful in a research project in which the EPDS is the only measure used or when it is used as a first-stage screening scale to identify possible depression. Another advantage of the EPDS in epidemiological studies is that it can elicit a high response rate (95–97%) when sent by post, especially if there had been previous contact with the research team and there is follow-up of women who do not initially respond (Murray & Carothers, 1990; Roy *et al*, 1993).

In the community, the EPDS is useful in the secondary prevention of postnatal depression by identifying the early onset of depressive symptoms. It can be administered by a trained health visitor, practice nurse or midwife at a postnatal or child health clinic or on a home visit. When using the EPDS in primary care settings as a component of a screening programme, the 9/10 cut-off may be over-inclusive, so a cut-off of 12/13 is often recommended. It should be remembered that the EPDS screens only for depression and also that women who score below the cut-off may none the less have depression.

The scale is best administered by a health professional who is familiar with mental health problems and has had training in their evidence base. A woman with a high score or an unexpectedly low score should be further assessed by a health professional and/or referred to a general practitioner, mental health nurse, psychologist or psychiatrist. A mother with profound depression might not grasp the meaning of the

items; others may wish to cover up their disability or fear stigma or the shame of not coping.

Comparison with other scales

Several studies have compared the performance of the EPDS and other depression questionnaires for use in the postnatal period.

In the UK (Wales), Harris *et al* (1989) compared the EPDS and the BDI in their abilities to identify subjects who had major depression according to DSM-III criteria (American Psychiatric Association, 1980). The sensitivity of the EPDS was 95% and its specificity 93%. They concluded that the performance of the BDI was markedly inferior in this application, with a sensitivity of 68% and specificity of 88%.

In a Canadian study, Lussier *et al* (1996) found a low concordance between the BDI and the EPDS. Their analysis revealed distinct response patterns belonging to divergent subgroups, suggesting that the two instruments were differently attuned to the various aspects of the presentation of postnatal depression.

Another Welsh study (Thompson *et al*, 1998) found that the EPDS was superior to the HAD scale in identifying RDC-defined depression and similar to the observer-rated Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960), which it also matched for sensitivity to change in mood over time.

An Australian group (Condon & Corkindale, 1997) found little agreement between the EPDS, the depression sub-scale of the HAD scale, the Zung Self-Rating Depression Scale (SDS; Zung, 1965) and the depression sub-scale of the Profile of Mood States (POMS; McNair & Lorr, 1964). They concluded that this poor level of agreement might reflect the different emphasis in the item content of the questionnaires.

In France, Guedeney *et al* (2000) compared the EPDS with the GHQ-28 and the Centre for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977), and suggested that the EPDS was better at identifying depression in postnatal women with anhedonic and anxious symptomatology, but less satisfactory for women with psychomotor retardation.

Using a sample of 50 Austrian postnatal women taking part in an international epidemiological study who scored 7 or more on the German version of the EPDS, Muzik *et al* (2000) compared the EPDS, the German version of the Zung SDS (with a clinical cut-off score of 50), and the depression and anxiety sub-scales of the Symptom Checklist-90-Revised (SCL-90-R; Derogatis & Cleary, 1977) with a cut-off *T*-score of 63). Diagnoses of depression and anxiety were made using the Structured Clinical Interview for DSM-III-R. The authors reported that the newly developed German version of the EPDS screened reliably for postnatal depression, and that further research was needed to create screening measures for post-partum anxiety disorders.

In the USA, Cheryl Beck (Beck & Gable, 2001) developed a new screening measure, the Post-partum Depression Screening Scale (PDSS). This new scale, developed from qualitative research, was compared with the EPDS and the BDI and validated against a DSM-IV diagnostic interview. She found that the PDSS yielded the highest combination of sensitivity (91%) and specificity (72%) of the three instruments and concluded that researchers and clinicians need to be aware of the differential sensitivity of depression instruments that are supposed to be measuring the same construct. It is also possible that the differences in performance may reflect differences in the population investigated.

We developed the EPDS because other depression measures available at that time were problematic for use with postnatal women living in the community. Other scales may be developed in the future, perhaps specific to a local population, particularly if it is from a diverse cultural background. However, we recommend the continued use of the EPDS because it is so widely used in different countries and therefore allows useful comparison between findings.

Other uses of the EPDS

The EPDS is not specific only to detecting depression in the puerperium. It can also be used to screen for depression in the following: pregnancy (Murray & Cox, 1990; Green & Murray, 1994; Evans *et al*, 2001; Josefsson *et al*, 2001); terminal illness (Lloyd-Williams *et al*, 2000); and fathers (Ballard *et al*, 1994; Areias *et al*, 1996b; Matthey *et al*, 2001). It has been used to assess dysphoria in adoptive parents (Gair, 1999). Cox *et al* (1996) validated the scale for use with non-postnatal women and it has also been validated for use with the mothers and fathers of toddlers (Thorpe, 1993). The scale can be administered by computer with adequate acceptability and performance (Glaze & Cox, 1991). The EPDS is not a measure of general psychiatric morbidity and will not detect other common perinatal disorders.