Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis


CRD summary
This review concluded that the Edinburgh Postnatal Depression Scale was a precise and validated test for identifying post-natal depression. There was insufficient evidence that use of identification strategies improved clinical outcomes. The conclusions on diagnostic accuracy appear reliable. The lower methodological quality and more limited evidence of the effectiveness review suggest that its conclusions should be treated with some caution.

Authors' objectives
To assess the validity (diagnostic accuracy), acceptability (not discussed in this abstract), clinical effectiveness and cost-effectiveness of methods used to identify post-natal depression in primary care.

Searching
The authors searched the following databases from inception to February 2007: MEDLINE, CINAHL, PsycINFO, EMBASE, Maternity and Infant Care, Cochrane Central Register of Controlled Trials (CENTRAL), DARE, Cochrane Database of Systematic Reviews, Science Citation Index, National Research Register, ReFeR (Research Findings Register), metaRegister of Controlled Trials, Health Services Research Projects in Progress, LILACS, Inside Conferences, and Dissertation Abstracts. No language restrictions were imposed. Search terms were reported. The authors also used forward citation searching, checked reference lists and contacted study authors and subject experts.

Study selection
Diagnostic accuracy: Studies of women in the prenatal or post-natal (up to one year) period that compared an identification test for depression with a reference standard (diagnostic interview) were eligible for inclusion. Studies could be cross-sectional, case-control, cohort or randomised controlled trials (RCTs). Most studies evaluated the Edinburgh Postnatal Depression Scale (EPDS) alone or in combination with other strategies. Studies were conducted in a variety of populations and settings, with EPDS cut-off points for depression ranging from 7 to 16. Outcomes were assessed at different time points across studies.

Clinical effectiveness: RCTs or controlled trials in any setting that compared the implementation of an identification strategy, with or without additional enhancement of care, with standard primary care were eligible for inclusion. Studies had to report maternal and/or infant outcomes. Included trials compared use of an identification strategy (followed by treatment of those identified) with usual care or used an identification strategy in all women and evaluated the effect of feeding back the results. These trials were designated level I or II evidence. Trials that used the identification strategy solely to assess eligibility (i.e. to exclude women without post-natal depression) were also eligible (level III evidence) and formed the majority of included studies.

Two reviewers independently selected studies for inclusion; disagreements were resolved by consensus, with reference to a third reviewer if necessary.

Assessment of study quality
Validity of diagnostic accuracy studies was assessed by two independent reviewers using the QUADAS (positive predictive value) checklist.

The authors did not state that they assessed validity of clinical effectiveness studies.

Data extraction
Diagnostic accuracy: Sensitivity, specificity, positive and negative likelihood ratios and diagnostic odds ratios were extracted or calculated. Data were extracted from English language articles by two independent reviewers. Articles in
other languages were extracted by one reviewer with the help of a translator.

**Clinical effectiveness**: For dichotomous outcomes, numbers of participants and outcomes in each group were extracted to calculate an odds ratio (OR) and associated 95% confidence interval (CI). For continuous outcomes, group means and standard deviations were extracted when reported to calculate a mean difference and associated 95% confidence interval. Data were extracted by one reviewer.

**Methods of synthesis**

**Diagnostic accuracy**: Studies that used the same identification strategy and cut point were pooled using a bivariate analysis. Pooled estimates of sensitivity, specificity, positive and negative likelihood ratios and diagnostic odds ratios and associated 95% confidence intervals were calculated. Between-study heterogeneity was assessed using the $I^2$ statistic of the pooled diagnostic odds ratio. To explore heterogeneity, summary receiver-operating characteristic (ROC) curves were constructed and used to identify studies that lay outside a 95% confidence ellipse in ROC space. A weighted multivariate meta-regression analysis was used to evaluate the influence of time since birth, baseline prevalence of depression and study quality features as sources of heterogeneity.

**Clinical effectiveness**: Pooled odds ratios for level I or II trials were calculated using a fixed-effect model (Mantel-Haenszel method). Statistical heterogeneity was assessed using the $I^2$ statistic. Continuous outcomes were not pooled because of reporting limitations. Trials providing level III evidence were synthesised in a narrative by type of study (prevention or treatment) and identification strategy used.

No evidence was found for the diagnostic accuracy or clinical effectiveness of strategies involving case finding questions.

**Results of the review**

**Diagnostic accuracy**: Sixty studies (64 publications) were included, of which 52 evaluated the Edinburgh Postnatal Depression Scale (EPDS). For the EPDS administered postnatally (47 studies), pooled estimates of sensitivity for major depression (18 studies) ranged from 0.60 (specificity 0.97) to 0.96 (specificity 0.45) depending on the cut point used. $I^2$ ranged from 0 to 93%. Quality varied between studies and some items were poorly reported. Generic identification strategies (Beck Depression Inventory and Hamilton Rating Scale for Depression) appeared to be less sensitive, but more specific, than the EPDS. Meta-regression analysis identified timing of administration of the EPDS (before or after six weeks postnatally) as a significant factor in explaining heterogeneity. Other analyses were reported.

**Clinical effectiveness**: Thirty studies (32 publications) were included. Five studies provided level I or II evidence. All the level I or II studies used the EPDS to identify post-natal depression. Pooled estimates (based on four trials) indicated that use of the EPDS significantly reduced the odds of subsequent post-natal depression, i.e. scoring above a defined threshold on the EPDS (OR 0.64, 95% CI 0.52 to 0.78). Heterogeneity was not significant ($I^2=0\%$). Three of the studies involved some form of enhanced care, making it difficult to isolate the effect of the identification strategy. Results for trials providing level III evidence were reported.

**Authors’ conclusions**

The Edinburgh Postnatal Depression Scale was a simple, safe, precise and validated test for the identification of postnatal depression. There was insufficient evidence to conclude that identification strategies were effective in improving maternal and infant outcomes.

**CRD commentary**

The diagnostic accuracy and clinical effectiveness reviews had clear and appropriate inclusion criteria. The authors searched a wide range of relevant sources, without language restrictions, and attempted to locate ongoing and unpublished studies. However, publication bias was not formally evaluated.

Validity was assessed appropriately in the diagnostic accuracy review, but it appeared that the authors did not assess validity in the clinical effectiveness review. This meant that the reliability of the included studies and the synthesis...
based on them was uncertain. Measures to minimise reviewer errors and bias were used in the diagnostic accuracy review, but again the clinical effectiveness review appeared to be less rigorous. Full details of included studies were reported. Meta-analysis was used selectively in both reviews, reflecting clinical and methodological differences between studies; heterogeneity was explored.

The conclusions of the diagnostic accuracy review are likely to be reliable. The findings of the effectiveness review also reflected the evidence presented, but its lower methodological quality and more limited evidence base suggest that the conclusions should be treated with some caution.

**Implications of the review for practice and research**

**Practice:** The authors did not state any implications for practice.

**Research:** The authors stated that research is required to compare the diagnostic performance of the EPDS, case finding questions and generic depression measures, and to evaluate the clinical effectiveness of the strategy identified as the most valid and acceptable.

**Funding**
NIHR Health Technology Assessment Programme, project number 05/39/06.

**Bibliographic details**

**PubMedID**
19624978

**Linked records**

- Screening for postnatal depression: is it acceptable to women and healthcare professionals? A systematic review and meta-synthesis

**DOI**
10.3310/hta13360

**Original Paper URL**
http://www.hta.ac.uk/project/1521.asp

**Other URL**
Link to record on HTA database: http://www.crd.york.ac.uk/crdweb/ShowRecord.asp?AccessionNumber=32007000331& UserID=0
Link to article on NHS EED: http://www.crd.york.ac.uk/erweb/ShowRecord.asp?AccessionNumber=2201000001& UserID=0

**Additional Data URL**

**Other publications of related interest**

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Adult; Depression, Postpartum /diagnosis; Diagnostic Techniques and Procedures; Evidence-Based Medicine; Female; Humans; Primary Health Care

**AccessionNumber**
12009109883

**Date bibliographic record published**
02/12/2009

**Date abstract record published**
14/04/2010

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.