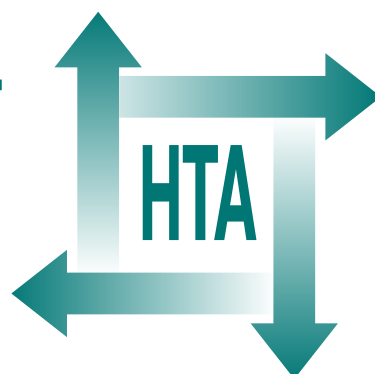# Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis
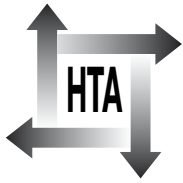
CE Hewitt, SM Gilbody, S Brealey,
M Paulden, S Palmer, R Mann,
J Green, J Morrell, M Barkham,
K Light and D Richards

**Health Technology Assessment
NIHR HTA programme
www.hta.ac.uk**

HTA

How to obtain copies of this and other HTA programme reports

An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch   Email: orders@hta.ac.uk
c/o Direct Mail Works Ltd   Tel: 02392 492 000
4 Oakwood Business Centre   Fax: 02392 478 555
Downley, HAVANT PO9 2NP, UK   Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can be purchased only for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA programme and lists the membership of the various committees.

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

CE Hewitt,[1] SM Gilbody,[1]* S Brealey,[1]
M Paulden,[2] S Palmer,[2] R Mann,[1] J Green,[3]
J Morrell,[4] M Barkham,[5] K Light[6]
and D Richards[1]

[1]Department of Health Sciences, University of York, UK
[2]Centre for Health Economics, University of York, UK
[3]Mother and Infant Research Unit, University of York, UK
[4]School of Health and Related Research, University of Sheffield, UK
[5]Department of Psychology, University of Sheffield, UK
[6]Centre for Reviews and Dissemination, University of York, UK

*Corresponding author

Declared competing interests of authors: none

NIHR Health Technology Assessment programme

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

Editor-in-Chief:	Professor Tom Walley CBE
Series Editors:	Dr Aileen Clarke, Dr Chris Hyde, Dr John Powell, Dr Rob Riemsma and Professor Ken Stein

Abstract

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

CE Hewitt,[1] SM Gilbody,[1*] S Brealey,[1] M Paulden,[2] S Palmer,[2] R Mann,[1] J Green,[3] J Morrell,[4] M Barkham,[5] K Light[6] and D Richards[1]

[1]Department of Health Sciences, University of York, UK
[2]Centre for Health Economics, University of York, UK
[3]Mother and Infant Research Unit, University of York, UK
[4]School of Health and Related Research, University of Sheffield, UK
[5]Department of Psychology, University of Sheffield, UK
[6]Centre for Reviews and Dissemination, University of York, UK
*Corresponding author

Objectives: To provide an overview of methods to identify postnatal depression (PND) in primary care and to assess their validity, acceptability, clinical effectiveness and cost-effectiveness, to model estimates of cost, to assess whether any method meets UK National Screening Committee (NSC) criteria and to identify areas for future research.
Data sources: Searches of 20 electronic databases (including MEDLINE, CINAHL, PsycINFO, EMBASE, CENTRAL, DARE and CDSR), forward citation searching, personal communication with authors and searching of reference lists.
Review methods: A generalised linear mixed model approach to the bivariate meta-analysis was undertaken for the validation review with quality assessment using QUADAS. Within the acceptability review, a textual narrative approach was employed to synthesise qualitative and quantitative research evidence. For the clinical and cost-effectiveness reviews methods outlined by the Centre for Reviews and Dissemination and the Cochrane Collaboration were followed. Probabilistic models were developed to estimate the costs associated with different identification strategies.
Results: The Edinburgh Postnatal Depression Scale (EPDS) was the most frequently explored instrument across all of the reviews. In terms of test performance, postnatally the EPDS performed reasonably well: sensitivity ranged from 0.60 (specificity 0.97) to 0.96 (specificity 0.45) for major depression only; from 0.31 (specificity 0.99) to 0.91 (specificity 0.67) for major or minor depression; and from 0.38 (specificity 0.99) to 0.86 (specificity 0.87) for any psychiatric disorder. Evidence from the acceptability review indicated that,

in the majority of studies, the EPDS was acceptable to women and health-care professionals when women were forewarned of the process, when the EPDS was administered in the home, with due attention to training, with empathetic skills of the health visitor and due consideration to positive responses to question 10 about self-harm. Suggestive evidence from the clinical effectiveness review indicated that use of the EPDS, compared with usual care, may lead to reductions in the number of women with depression scores above a threshold. In the absence of existing cost-effectiveness studies of PND identification strategies, a decision-analytic model was developed. The results of the base-case analysis suggested that use of formal identification strategies did not appear to represent value for money, based on conventional thresholds of cost-effectiveness used in the NHS. However, the scenarios considered demonstrated that this conclusion was primarily driven by the costs of false positives assumed in the base-case model.
Conclusions: In light of the results of our evidence synthesis and decision modelling we revisited the examination of PND screening against five of the NSC criteria. We found that the accepted criteria for a PND screening programme were not currently met. The evidence suggested that there is a simple, safe, precise and validated screening test, in principle a suitable cut-off level could be defined and that the test is acceptable to the population. Evidence surrounding clinical and cost-effectiveness of methods to identify PND is lacking. Further research should aim to identify the optimal identification strategy, in terms of key psychometric properties for postnatal populations. In particular,

research comparing the performance of the Whooley and help questions, the EPDS and a generic depression measure would be informative. It would also be informative to identify the natural history of PND over time and to identify the clinical effectiveness of the most valid and acceptable method to identify PND. Further research within a randomised controlled trial would provide robust estimates of the clinical effectiveness.

Contents

List of abbreviations

BDI	Beck Depression Inventory		EPDS	Edinburgh Postnatal Depression Scale
BPDS	Bromley Postnatal Depression Scale		EVPI	expected value of perfect information
CCEI	Crown–Crisp Experimental Index		EVPPI	expected vale of partial perfect information
CEAC(s)	cost-effectiveness acceptability curve(s)		GHQ	General Health Questionnaire
CES-D	Center for Epidemiologic Studies Depression Scale		HADS	Hospital Anxiety and Depression Scale
CI	confidence interval		HAMA	Hamilton Rating Scale for Anxiety
CIDI	Composite International Diagnostic Interview		HAMD	Hamilton Rating Scale for Depression
CIS	Clinical Interview Schedule		HDRS	Hamilton Depression Rating Scale
CRD	Centre for Reviews and Dissemination		HRQoL	health-related quality of life
DAS	Dyadic Adjustment Scale		HRSD	Hamilton Rating Scale for Depression
DASS	Depression Anxiety Stress Scale		HSCL	Hopkins Symptom Checklist
DIS	Diagnostic Interview Schedule		HSROC	hierarchical summary receiver operating characteristic
DOR(s)	diagnostic odds ratio(s)		HTA	Health Technology Assessment
DSM	*Diagnostic and Statistical Manual of Mental Disorders*			

ICD	*International Classification of Diseases, Injuries and Causes of Death*		PCT(s)	primary care trust(s)
			PDS	Pitt Depression Scale
ICER	incremental cost-effectiveness ratio		PDSS	Postpartum Depression Screening Scale
LR+	Positive Likelihood Ratio		PHQ	Patient Health Questionnaire
LR−	Negative Likelihood Ratio		PI	Predictive Index
MADRS	Montgomery–Asberg Depression Rating Scale		PND	postnatal depression
			POMS	Profile of Mood States
MASQ	Modified Antenatal Screening Questionnaire		PRQ	Pregnancy Risk Questionnaire
MINI	Mini-International Neuropsychiatric Interview		PRIME-MD	Primary Care Evaluation of Mental Disorders
MOS	Medical Outcomes Study depression measure		PSE	Present State Examination
MSSS	Maternity Social Support Scale		PSS	personal social services
			QALY	quality-adjusted life-year
NHS EED	NHS Economic Evaluations Database		QUADAS	Quality Assessment of Diagnostic Accuracy Studies
NICE	National Institute for Health and Clinical Excellence		RCT(s)	randomised controlled trial(s)
NRR	National Research Register		RDC	Research Diagnostic Criteria
NSC	National Screening Committee		RDRS	Raskin Depression Rating Scale
OR	odds ratio		ROC	Receiving Operating Characteristic
PAS	Psychiatric Assessment Schedule			

SADS	Schedule for Affective Disorders and Schizophrenia	SE	standard error
SCID	Structured Clinical Interview for DSM	SPI	Standardised Psychiatric Interview
SCL-90-R	Symptom Checklist-90-Revised	sROC	summary receiver operating characteristic
SDSS-PC	Symptom-Driven and Diagnostic System for Primary Care	VOI	value of information
		WTP	willingness to pay
SDS	Zung's Self-rating Depression Scale	Zung's SDS	Zung's Self-rating Depression Scale

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or in the notes at the end of the table.

ix

Executive summary

Background

Depression accounts for the greatest burden of disease among all mental health problems, and is expected to become the second-highest among all general health problems by 2020. Postnatal depression (PND) is an important category of depression in its own right. There is now considerable evidence to show that PND has a substantial impact on the mother and her partner, the family, mother–baby interactions and the longer-term emotional and cognitive development of the baby, especially when depression occurs in the first year of life. Unfortunately, less than 50% of cases of PND are identified by primary health-care professionals in routine clinical practice. PND screening and case identification strategies have been advocated as a remedy to this problem, but this has attracted substantial controversy.

Objectives

1. To provide an overview of all available methods to identify PND and to assess their validity (in terms of key psychometric properties).
2. To assess the acceptability of methods to identify PND.
3. To assess the clinical effectiveness of methods to identify PND in improving maternal and infant outcomes.
4. To assess the cost-effectiveness of methods to identify PND in improving maternal and infant outcomes.
5. To identify research priorities and the value of further research into methods to identify PND from the perspective of the UK NHS.
6. To assess whether methods to identify PND meet minimum criteria outlined by the National Screening Committee (NSC) in the light of this evidence synthesis.

Methods

A large search was undertaken across all phases of the review, which involved searching 20 electronic databases (including MEDLINE, CINAHL, PsycINFO, EMBASE, CENTRAL, DARE and CDSR), forward citation searching of key literature, personal communication with authors and scrutinising reference lists. A variety of review methods were utilised across the four systematic reviews. A generalised linear mixed model approach to the bivariate meta-analysis was undertaken for the validation review with quality assessment using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. Within the acceptability review, a textual narrative approach was employed to synthesise qualitative and quantitative research evidence. For the clinical and cost-effectiveness reviews, methods outlined by the Centre for Reviews and Dissemination and the Cochrane Collaboration were followed. Probabilistic models were developed to estimate the costs associated with different identification strategies. Scenario-based sensitivity analyses were also performed.

Results

There were numerous generic and PND-specific measures identified that may be used to identify possible cases of PND. A total of 14 identification strategies were found to have been validated among women during pregnancy or the postnatal period: PND-specific measures that were used were the Edinburgh Postnatal Depression Scale (EPDS), Postpartum Depression Screening Scale, Pregnancy Risk Questionnaire, and Predictive Index; generic depression identification strategies were the Beck Depression Inventory (BDI), General Health Questionnaire (GHQ), Hospital Anxiety and Depression Scale, Hopkins Symptom Checklist, Hamilton Rating Scale for Depression (HAMD), Zung's Self-rating Depression Scale, Symptom Checklist-90-R, Raskin, and Montgomery–Asberg Depression Rating Scale; one study used both the EPDS and GHQ. By far the most frequently used identification strategy across all of the reviews was the EPDS. In terms of test performance, postnatally the EPDS performed reasonably well: sensitivity ranged from 0.60 (specificity 0.97) to 0.96 (specificity 0.45) for major depression only; from 0.38 (specificity 0.99) to 0.86 (specificity 0.87) for any psychiatric disorder; and from 0.31 (specificity 0.99) to 0.91 (specificity 0.67) for major

or minor depression. In addition, for major or minor depression there were sufficient data to pool the BDI and HAMD data at a single cut point. Results from this analysis highlighted that generic identification strategies may be less sensitive than the EPDS, but more specific.

For the acceptability review, studies indicated that women and health professionals both felt that it was beneficial to inform women in advance that they would be asked to complete a questionnaire to identify PND and that the questionnaire should be administered in the woman's home. In general, when administering the instrument, women preferred to talk rather than complete a standardised questionnaire and were critical of the lack of dialogue that could result from a paper and pencil assessment. Both women and health professionals found that the last question on the EPDS, about the thought of self harm, caused difficulties. In addition, English women and health professionals also found difficulties with the question about sleeping. It was also identified that the interpersonal relationship between the mother and health professional was important and that this relationship was strengthened after a number of meetings and when adequate training for health professional in identifying PND was given. In summary, in the majority of studies, the EPDS was acceptable to women and healthcare professionals when women were forewarned of the process, when the EPDS was administered in the home, with due attention to training those administering the EPDS, with empathetic skills of the health visitor and due consideration of positive responses to question 10 about self harm.

Within the clinical effectiveness review, five studies were identified that compared using either the EPDS (with or without enhancement of care) or feedback of the EPDS scores with not using the EPDS or usual care. All of the studies indicated beneficial effects of using the EPDS in reducing EPDS scores, although some of the individual studies did not show statistically significant differences. Studies reporting dichotomous outcomes (the number of women scoring above or below a cut point on the EPDS) were combined and the pooled estimate gave an odds ratio of 0.64 (95% confidence interval 0.52 to 0.78). It was difficult to disentangle the effects of using an identification strategy from the effects of the enhancement of care and/or any subsequent intervention given.

With regards to the cost-effectiveness of methods to identify PND, despite an extensive systematic search of the literature, none of the studies identified presented full economic evaluations of PND identification strategies, hence a decision-analytic model was developed. The results of the base-case analysis suggested that the use of formal identification strategies did not appear to represent value for money based on conventional thresholds of cost-effectiveness used in the NHS. However, the scenarios considered demonstrated that this conclusion was primarily driven by the costs of false positives assumed in the base-case model. Alternative assumptions employed in separate scenarios resulted in more favourable estimates of cost-effectiveness, such that use of the EPDS to identify women with PND, considered in some of these scenarios, fell within these conventional thresholds. For example, when the cost of a false-positive diagnosis was assumed to be a single GP attendance, the EPDS using a cut point of 10 or higher emerged as the optimal strategy in terms of cost-effectiveness. Interestingly, this corresponded closely with the results presented in the validation review, in which the trade-off between sensitivity and specificity was considered. A definitive answer to the question of whether formal identification strategies are cost-effective, and, if they are, which individual strategy is optimal in cost-effectiveness terms, clearly requires more reliable evidence in relation to the costs of managing false positives.

Clinical guidance on the management of antenatal and postnatal mental health care was issued by the National Institute for Health and Clinical Excellence (NICE) in October 2007. NICE recommended the use of the Whooley questions:

1. 'During the past month, have you often been bothered by feeling down, depressed or hopeless?'
2. 'During the past month, have you often been bothered by little interest or pleasure in doing things?'

A third help question should be considered if the woman answers 'yes' to either of the initial questions:

3. 'Is this something you feel you need or want help with?'

No evidence was identified across the four systematic reviews for these three questions in a postnatal population in terms of validity, acceptability and clinical and cost-effectiveness.

Conclusions

In light of the results of our evidence synthesis and decision modelling we revisited the examination of PND screening against five of the NSC criteria. We found that the accepted criteria for a PND screening programme were not currently met. The evidence suggested that there is a simple, safe, precise and validated identification strategy, that in principle a suitable cut-off level could be defined and that the strategy is acceptable to the population. Evidence surrounding the clinical effectiveness and cost-effectiveness of methods to identify PND is lacking.

Implications for research

The results from the systematic reviews, the probabilistic decision model and the value of information analysis indicated that further research should aim to identify the:

- Optimal identification strategy, in terms of key psychometric properties, for postnatal populations. Further research comparing the performance of the Whooley and help questions, the EPDS and a generic depression measure would be informative.
- Acceptability of the identification strategies outlined above, with particular emphasis on collating acceptability data by whether women were correctly classified (i.e. true positives or true negatives) or not (i.e. false positives or false negatives).
- Natural history of PND over time in populations in which formal methods to identify PND have been used and in populations in which formal methods of identification have not been used.
- Costs associated with false positives.
- Impact of PND on health-related quality of life.
- Epidemiological data regarding prevalence rates of PND.
- Clinical effectiveness of the most valid and acceptable method to identify PND. This could be achieved by carrying out further research within a randomised controlled trial.

Chapter 1
Introduction and background

Depression accounts for the greatest burden of disease among all mental health problems, and is expected to become the second-highest among all general health problems by 2020.[1] Postnatal depression (PND) is an important category of depression in its own right. There is now considerable evidence to show that postnatal depression has a substantial impact on the mother and her partner,[2] the family,[3] mother–baby interactions[4] and the longer-term emotional and cognitive development of the baby,[5] especially when depression occurs in the first year of life.[6] Unfortunately, less than 50% of cases of PND are detected by primary health-care professionals in routine clinical practice.[7] PND screening and case identification strategies have been advocated as a remedy to this problem, but this has attracted substantial controversy.[8] In 2006 the Health Technology Assessment (HTA) programme prioritised a review of the clinical validity, utility and cost-effectiveness of methods for the identification of PND in primary care and the results are presented in this report.

Definitions and epidemiology of postnatal depression

For the purposes of this research, PND was defined as a non-psychotic depressive episode meeting standardised diagnostic criteria for a minor or major depressive disorder, beginning in or extending into the postnatal period.[9] Within our research we distinguish postnatal depression from other types of mental health problems that can also occur during pregnancy and the postnatal period. Baby blues and puerperal psychosis are two such examples and are important health problems within their own rights. Baby blues and puerperal psychosis were not addressed specifically within this body of research. From a clinical perspective, PND includes three subgroups of women whose management may differ: (1) those who develop depression only after childbirth; (2) those who have developed antenatal depression, which continues into the postnatal period; and (3) those women with pre-existing chronic or relapsing depression. A meta-analysis of 59 studies (including 12,810 women, mainly from first world countries) found

that the prevalence of depression within the first few postnatal months was 13% [95% confidence interval (CI) 12.3% to 13.4%].[10] Most cases develop within the first 3 postnatal months,[11] with a peak incidence at around 4–6 weeks.[9] Although one study[11] showed that most cases last around 3 months and resolve spontaneously without treatment, another study[12] demonstrated the presence of depression with over 50% of cases lasting over 6 months and some being still present at 4 years.

Clinical and social consequences of postnatal depression

Depression in all populations is associated with profound decrements in quality of life, social functioning and economic productivity.[13] However, in the case of PND, the adverse consequences are felt beyond the individual with depression, affecting the family and development of the infant. In particular, the severity and chronicity of maternal depression are predictive of disturbances in child development.[14] Physical health and risk of childhood injury also seem to be adversely affected as a consequence of PND.[15]

Strategies to improve the detection of postnatal depression

Given the adverse consequences of PND and the general underidentification of this problem, a number of strategies have been proposed to improve PND identification. These broadly fall into five categories:

- postnatal identification using specially developed standardised postnatal questionnaires [such as the Edinburgh Postnatal Depression Scale (EPDS)[16] and the US Postpartum Depression Screening Scale (PDSS)[17]
- postnatal identification using standardised generic questionnaires for depression [such as the Beck Depression Inventory (BDI)[18]

- prenatal screening using standardised depression questionnaires to identify those with pre-existing depression and those at risk of developing significant depression in the postnatal period[19]
- prenatal screening using known risk factors for PND (such as previous history of depression and lack of social support) to identify those who are likely to subsequently develop depression in the postnatal period[20]
- the use of training packages targeted at health-care professionals designed to enhance awareness and recognition of clinical signs of PND and to ensure that a thorough psychosocial assessment is provided on a routine basis.[21]

Current policy and practice within the UK

Following publication of authoritative recommendations in a number of national policy documents,[22,23] the use of case-finding and screening strategies has accelerated rapidly. For example, the National Service Framework made an explicit requirement that all areas should have local protocols for the management of PND.[23] In practice, screening and case-finding strategies have dominated and have tended to focus upon the routine or ad hoc administration of the EPDS in the postnatal period, such that it has become the most widely used of the above strategies. The application of this measure has tended to fall upon health visitors, who are responsible for monitoring the well-being of the mother and newborn after 14–28 days postnatally. The de facto implementation of a national screening strategy in the UK, based around the EPDS, has attracted substantial discussion. Criticisms have been levelled at this strategy based upon the ethics of mass screening; the psychometric properties of available instruments (especially the EPDS); the acceptability (to both patients and health-care professionals) of screening and case-finding strategies; and the absence of any evidence that screening, per se, leads to effective management and improved mother and infant outcomes.[8]

Screening is only one way in which recognition and management of PND might be improved and there are clear criteria laid down that should guide the adoption of a screening strategy as part of national health policy. Screening tests can be justified only if the instrument is accurate, results in a more

effective treatment than would otherwise be the case and does so with a favourable ratio of costs to benefits.[24] These criteria have been codified in the UK by the establishment of the National Screening Committee (NSC) and the publication of clear criteria that must be satisfied before adoption of a screening strategy.[25] When these criteria were applied to the case of screening for PND,[26] insufficient evidence was found to support the implementation of this strategy. This NSC recommendation has not been without controversy, but genuine concerns remain regarding the acceptability, validly, clinical effectiveness and cost-effectiveness of identification methods for PND.

Several reviews have been undertaken to identify, review and assess the performance of methods to identify PND.[27–29] The most recent of these is a comprehensive review of antenatal and postnatal care that was undertaken by the National Institute for Health and Clinical Excellence (NICE) in 2007.[30] On the basis of this review, clinical guidance on the management of antenatal and postnatal mental health care was issued. The guidance included a review of methods to identify mental health problems during the postnatal period. The review considered two methods of identification, the EPDS and case-finding questions (Whooley plus help question).

A literature search was undertaken to investigate the psychometric properties of the EPDS. Eight validation studies[16,31–37] and a recent systematic review[27] were identified from the literature; the evidence from the systematic review[27] alone was used to establish the diagnostic performance of the EPDS. In the review three studies were retrieved that had used the EPDS to identify major depression during the postnatal period. Pooling of two of these studies was undertaken at a cut point of ≥ 13 and the EPDS was found to have a sensitivity of 0.91 (95% CI 0.84 to 0.99) and a specificity of 0.91 (95% CI 0.88 to 0.94). Based on these figures, together with a prevalence rate of 6.8% (based on a review of prevalence studies performed by the same authors), an overall positive predictive value of 57% and a negative predictive value of 99% was calculated. Furthermore, pooling of three studies was undertaken at a cut point of ≥ 10 to identify women with major or minor depression and the EPDS was found to have a sensitivity of 0.68 (95% CI 0.58 to 0.78); pooling of specificity values was not undertaken because of significant heterogeneity although a value of 0.80 is reported in the discussion. Based on these figures,

together with a prevalence rate of 11.3% (based on a review of prevalence studies performed by the same authors), an overall positive predictive value of 30% and a negative predictive value of 95% were calculated.

Evidence of the psychometric properties of case-finding questions came from two studies.[38,39] In the first study[38] the potential advantages of two brief questions (termed 'Whooley questions') compared with usual measures for identifying depression (depression in general not PND) were explored. The two brief questions were derived from the 2-item Patient Health Questionnaire (PHQ)-9 and were: (1) 'During the past month, have you often been bothered by feeling down, depressed or hopeless?' and (2) 'During the past month, have you often been bothered by little interest or pleasure in doing things?' The two questions, Center for Epidemiologic Studies Depression Scale (CES-D), Medical Outcomes Study depression measure (MOS), BDI, Symptom-Driven Diagnostic System for Primary Care (SDDS-PC) and Diagnostic Interview Schedule (DIS) were administered to male participants attending a Veterans Affairs Medical Center in San Francisco. A positive response to the two case-finding questions had a sensitivity of 0.96 (95% CI 0.90 to 0.99) and a specificity of 0.57 (95% CI 0.53 to 0.62). Based on these figures, together with a prevalence rate for major depression of 18.1%, an overall positive predictive value of 33% and a negative predictive value of 98% were calculated.

The second study[39] used the same questions as the previous study with an additional 'help' question: 'Is this something with which you would like help?' The three questions (termed 'Whooley questions plus the help question') were administered to 936 patients through 19 general practitioners in six clinics in New Zealand and were validated against the standardised psychiatric interview CIDI (Composite International Diagnostic Interview). A positive response to either question plus the 'help' question had a sensitivity of 0.96 (95% CI 0.86 to 0.99) and a specificity of 0.89 (95% CI 0.87 to 0.91) when identifying major depression. Based on these figures, together with a prevalence rate for major depression of 5%, an overall positive predictive

value of 32% and a negative predictive value of 99.7% were calculated.

The current NICE guidance issued on antenatal and postnatal mental health recommends the use of the Whooley questions plus the additional help question and states:

- At a woman's first contact with primary care, and at her booking visit and postnatally (usually at 4–6 weeks and 3–4 months), health-care professionals (including midwives, obstetricians, health visitors and GPs) should ask the two questions to identify possible depression:
  – 'During the past month, have you often been bothered by feeling down, depressed or hopeless?'
  – 'During the past month, have you often been bothered by little interest or pleasure in doing things?'
- A third question should be considered if the woman answers 'yes' to either of the initial questions:
  – 'Is this something you feel you need or want help with?'
- Health-care professionals may consider the use of self-report measures such as the EPDS, Hospital Anxiety and Depression Scale (HADS) or PHQ-9 as part of subsequent assessment or for the routine monitworing of outcomes.

The Whooley case-finding questions, derived from the Primary Care Evaluation of Mental Disorders (PRIME-MD), are brief and easy to use in routine practice and are currently recommended in the NICE guidance as the identification method of choice for case finding minor and major depression in any type of non-postnatal population. However, no research literature currently exists of studies that have considered, or have validated, case-finding questions in samples of women in the postnatal period. The NICE research recommendations based on the review propose that a validation study should be undertaken of the Whooley questions plus the help question in women in the first postnatal year, examining the effectiveness of the questions when used by midwives and health visitors compared with a psychiatric interview.

Chapter 2
Aims and objectives

The purpose of this research was to apply rigorous systematic review and evidence synthesis techniques to evaluate methods to identify PND. There were several objectives:

1. to provide an overview of all available methods to identify PND in primary care and to assess their validity (in terms of key psychometric properties)
2. to assess the acceptability of methods to identify PND in primary care
3. to assess the clinical effectiveness of methods to identify PND in improving maternal and infant outcomes in primary care
4. to assess the cost-effectiveness of methods to identify PND in improving maternal and infant outcomes in primary care
5. to identify research priorities and the value of further research into methods to identify PND from the perspective of the UK NHS
6. to assess whether methods to identify PND meet minimum criteria outlined by the NSC in the light of this evidence synthesis.

Chapter 3

Literature searching

We summarised and synthesised the available research literature regarding identification strategies for PND. A range of study designs (qualitative, quantitative, descriptive and economic) were synthesised, relevant to each of the specific aims outlined in Chapter 2. At all phases of the review we adhered to accepted guidelines outlined by the Centre for Reviews and Dissemination (CRD),[40] with specific adaptations to reflect innovations in synthesising psychometric[41] and economic data,[42] and in prioritising research.[43] One large search was undertaken across all of the phases of the review, rather than individual searches for each review.

Search strategy

Searches were undertaken on the following databases to identify relevant clinical and cost-effectiveness literature:

- MEDLINE (Ovid Online – www.ovid.com/)
- CINAHL (Cumulative Index to Nursing and Allied Health Literature) (Ovid Online – www.ovid.com/)
- PsycINFO (Ovid Online – www.ovid.com/)
- EMBASE (Ovid Online – www.ovid.com/)
- Maternity and Infant Care (Ovid Online – www.ovid.com/)
- CENTRAL and DARE (Database of Abstracts of Reviews of Effects) (Cochrane Library – CD-ROM)
- CDSR (Cochrane Database of Systematic Reviews) (Cochrane Library – CD-ROM)
- Science Citation Index (SSCI) (Web Of Knowledge – http://wos.mimas.ac.uk/)
- NRR (National Research Register) (www.nrr.nhs.uk/)
- ReFeR (Research Findings Register)
- *meta*Register of Controlled Trials (*m*RCT) (via Current Controlled Trials – http://controlled-trials.com/)
- Health Services Research Projects in Progress (HSRProj) (www.nlm.nih.gov/hsrproj/)
- LILACS (http://bases.bireme.br/cgibin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&base=LILACS&lang=I)

- Inside Conferences – accessed via Dialog (file 65) using DialogLink 5
- Dissertation Abstracts – accessed via Dialog (file 35) using DialogLink 5.

In addition, the following databases were searched specifically for cost-effectiveness studies:

- NHS Economic Evaluation Database (NHS EED) (CRD – www.york.ac.uk/inst/crd/crddatabases.htm)
- Health Economic Evaluations Database (HEED) (CD-ROM)
- IDEAS (http://ideas.repec.org/)
- EconLit (ERLWebSPIRS5 – http://arc.uk.ovid.com/).

Forward citation searching

For the validation review (Chapter 5), forward citation searching was undertaken for the original EPDS[16] and PDSS[44] validation studies. This process was undertaken using the Web of Science (WoS) Institute of Scientific Information (ISI) citation database. Each study was entered separately and all citations to the paper since publication were identified. Titles and, when available, abstracts of the papers that had cited the selected trials were downloaded.

Terminology

The terms for the search strategies were identified through discussion between an information officer and the research team, by scanning the background literature and by browsing the MEDLINE thesaurus (MeSH). All databases were searched from their inception to the date of the search. Searches took place during February 2007 (see Appendix 1 for dates of individual searches). No language or other restrictions were applied. A broader strategy was used on the economic databases to capture primary economic evaluations relating to depression. Full details of the search strategies are reported in Appendix 1.

Inclusion and exclusion criteria

All records were imported into a bibliographic referencing software program (ENDNOTE version 9). Duplicate references were identified using the inbuilt function within ENDNOTE and subsequently deleted. Two reviewers screened titles and abstracts to identify potentially eligible studies; any disagreements were resolved by consensus or deferred to a third party if necessary. Studies were assessed for inclusion across all phases of the review. Further details regarding the inclusion and exclusion criteria for each individual review can be found in the following chapters. Full papers of potentially eligible studies were obtained and assessed for inclusion independently by two reviewers.

Summary of literature searching

The study selection process is outlined in *Figure 1*. In total, 11,945 potentially relevant studies were identified from the searches, of which 225 were selected for full assessment. Of these, 108 studies met the inclusion criteria and were included in one or more of the reviews outlined in the following chapters.

Reference lists of all reviews were inspected to ensure that all potentially relevant studies had been identified. Authors were contacted when studies were published as abstracts or when there was insufficient information to assess eligibility or extract the relevant data. In addition, authors of ongoing and recently completed research projects were contacted to enquire if the research had been completed and if there were any subsequent publications. In total, 32 authors were contacted for further information and 20 authors responded providing further information.

Stakeholder involvement

Postnatal depression is an important health problem and PND identification has proved to be an especially controversial area of practice and policy. Our research sought to engage with, rather than ignore, this area of controversy. Important stakeholders were engaged in helping us frame our research questions and in understanding the results of our evidence synthesis.

Stakeholders can be thought of as groups of individuals who have specific interests and concerns with respect to a particular issue. Stakeholders are not a homogeneous group and there may be important differences between stakeholders in terms of their understanding of the issue and expectations of the actions that should be taken by others (e.g. regulators, policy-makers, professionals, members of the public). Stakeholder engagement requires an explicit analysis of the 'power' and 'stake' that are inherent in different stakeholder constituencies. A meaningful engagement of stakeholders requires that the following are
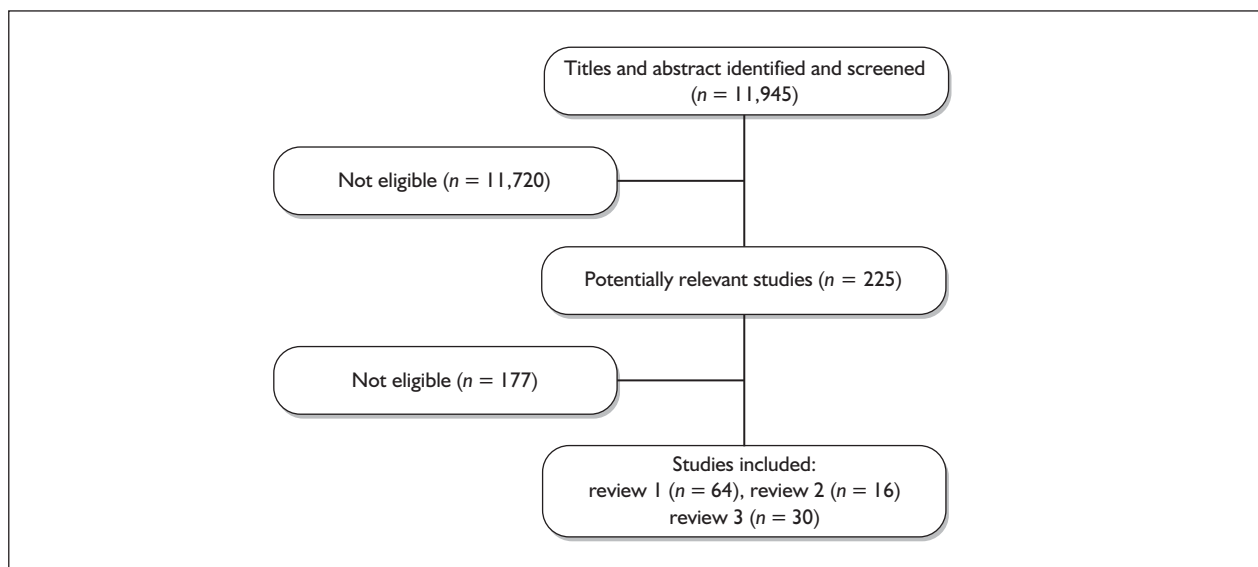
FIGURE 1 *Flow diagram of literature searches.*

included: (1) high stake/influence and high power (in this case Department of Health policy-makers; members of the NSC; commissioners of primary care and maternity services; national professional organisations – Community Practitioners and Health Visitors Association, Royal College of Psychiatrists, Royal College of Midwives) and (2) high stake, low influence [in this case frontline primary health-care workers – midwives, GPs, psychiatrists and health visitors – and women (and their partners) with experience of maternity services and PND].

We held a single stakeholder consultation day in November 2007. At the beginning of the consultation day, participants were provided with an overview of the background to PND from an epidemiological perspective, and to the principles of screening, with illustrations of the main screening, research, policy and practice issues for PND. Subsequently, stakeholders were asked to participate in the first focus group, which examined stakeholders:

- general perceptions of PND and screening
- awareness of methods to identify PND
- methods used in practice
- awareness of the recent NICE guidelines

- views about the Whooley + help case finding questions
- views about the EPDS.

The programme of research was then outlined and the emerging results presented. Following the presentations, the second focus group was undertaken, which examined:

- stakeholders views of the emerging research evidence presented
- influence of research evidence on the stakeholders prior beliefs
- importance of validity, acceptability, outcome and costs in defining good measures
- opinions on the predefined themes for the acceptability review
- research priorities.

Stakeholders were involved in the design, refinement, conduct and analysis of this programme of reviews. Health professionals, service users and researchers have contributed to and commented on the research at key stages in its development. A full list of the invited and attending stakeholders can be found in Appendix 2.

9

Chapter 4
Survey of available methods to identify postnatal depression

Introduction

To produce a comprehensive list of potential PND identification methods we undertook a scoping literature review. This review contextualises the systematic reviews presented in the following chapters. It provides an overview of the different classification systems, the criteria for diagnosis of major depression and the differences between the classification systems, the criteria for diagnosis of PND, and the different generic and postnatal-specific measures that are available.

Historical background of classification systems

Formal classification systems of mental health problems came to the forefront of psychiatry during the mid-twentieth century because of the need to provide a consistent and standardised approach to the classification of the heterogeneous range of symptoms associated with mental health problems such as depressive syndrome.[45] The aim of classification systems is to promote increased physician communication, understanding and consensus of diagnosis, enhanced understanding of the distinct differences between disorders such as unipolar and bipolar conditions, and therefore ultimately select the most effective and appropriate treatment available for a specific condition.[45]

The *Diagnostic and Statistical Manual of Mental Disorders*, 1st edition (DSM-I)[46] published in 1952 by the American Psychiatric Association and the addition of the mental disorders section to the *International Classification of Diseases, Injuries and Causes of Death, 6th revision* (ICD-6)[47] in 1948 by the World Health Organization represented the first attempts to achieve these aims. The DSM-I focused on the concept of 'reactive' aspects of psychiatric conditions. These were limited to the classification of disturbances of mental functioning and were designated as groups of related psychiatric syndromes, termed as disorders. The ICD-6 mental disorders section was limited to psychosis, psychoneurosis and disorders of character,

behaviour and intelligence. Lack of international acceptance of these classification systems and a shift in the concept of psychiatric nomenclature led to subsequent revisions of both systems, which, through close collaboration, resulted in substantial similarity between the DSM-II[48] and ICD-8.[49] By the mid-1970s, a crucial problem was highlighted with the DSM-II. There was a lack of explicit criteria for diagnosis, whereby a diagnostic category was selected based on whichever one most closely resembled the characteristics of the patient. The perceived need for consistent sets of criteria for clinical work and selection of research samples, and concerns that the diagnostic approaches in the development of earlier classification systems lacked reliability, led to the development of the Research Diagnostic Criteria (RDC).[50] The RDC, which elaborated on the earlier diagnostic work of the Feighner criteria,[51] was based on the concept that for each disorder explicit inclusion and exclusion criteria should exist based on symptoms or signs of illness, the level of severity or impairment experienced and the duration or course of illness.

The concepts of the RDC formed the basis for specified diagnostic criteria for all of the mental disorders included in the development of the DSM-III.[52] In contrast to the earlier editions, the DSM-III made the distinction between a diagnosis of major depressive episode and bipolar disorder, and distinguished between the presence and absence of mania. In addition, diagnoses of depressive 'reaction' and neurotic depression were withdrawn. The DSM-III, published in 1980, was adopted 1 year later as the official classification system within mental health facilities in the USA. The DSM-III represented a major shift from the ICD-9,[53] which did not contain explicit criteria until the development of the ICD-9-Clinical Modification[54] (ICD-9-CM), which provided a glossary of descriptions of abnormal mental behaviour that represented a consistent framework of reference. Subsequent revisions – DSM-IV,[55] ICD-10[56] – and the RDC form the current diagnostic framework for classification of mental disorders, representing the most widely accepted reference case or 'gold

standard' diagnostic procedure for establishing a psychiatric diagnosis.

The DSM-IV is a multiaxial system that organises each psychiatric diagnosis into five independent levels: axis I – major mood disorders including depression, anxiety and bipolar disorder; axis II – personality disorders; axis III – relevant general medical conditions and physical disorders; axis IV – relevant psychosocial and environmental stressors; and axis V – global assessment of functioning. DSM-IV reflects a global evaluation in which a person may be diagnosed with a disorder on more than one independent axis. The ICD-10 contains 22 chapters each with multilevel categories. Various criticisms have been directed at both classification systems for the unnecessary complexity of the categories, arbitrary and unvalidated boundaries between categories, lack of clarity in the precise meaning of manic states and the failure to give recognition to the close relationship between anxiety and depression.[57] Despite this the DSM-IV and ICD-10 numeric diagnostic codes associated with each mental disorder are considered appropriate methods for the collection and dissemination of psychiatric morbidity and mortality data throughout the world. Nevertheless, the codes are rarely utilised outside of research, health service administrative and insurance purposes (e.g. in the USA).[57]

Diagnosis and classification of depressive disorder

In the diagnosis and classification of mood disorders such as depression the RDC, DSM-IV and ICD-10 display many similarities, which is unsurprising given the development and use of the RDC, the short timescale between development of subsequent editions of each system and collaboration between the global psychiatry communities. Eight key symptoms are common between the three classification systems for a depressive episode: depressed mood, loss of interest or pleasure, disturbed sleep, altered appetite, decreased energy, inability to concentrate, psychomotor agitation or retardation and thoughts of suicide or death.

The DSM-IV uses the term 'major depressive episode' and the ICD-10 the term 'depressive episode' and neither attributes a clear aetiology to underlying biochemical processes. Structurally the systems differ. In the ICD-10 two sets of items are presented, described as 'typical' and 'common'

symptoms; one set contains three typical symptoms and the other seven common symptoms. In addition, a third set presents somatic symptoms. Mild, moderate or severe episodes are based on separate diagnostic thresholds dependent on the number, type and severity of the symptoms presented.

In the DSM-IV nine items are presented in one set and severity assigned after the criteria for major depressive diagnosis have been met. In contrast the RDC allows for cases to be defined as either 'probable' or 'definite', and requires a patient to have experienced impairment in daily activities of living, with help maybe being required, and explicit exclusion of schizophrenia for a diagnosis of major depression to be met.

Table 1 displays the diagnostic criteria for the classification of depressive disorder for the DSM-IV, ICD-10 and RDC and highlights the differences between the systems in terms of the type, duration and number of symptoms required and the method of classification of severity of depressive episode. The main difficulty with the differences between the systems is that potentially it may result in an individual being classified differently on the basis of severity or by recurrence, dependent on which diagnostic criteria are applied.

Diagnosis and classification of postnatal depression

Postnatal depression is classified in the ICD-10 under the category 'Behavioural syndromes associated with physiological disturbance and physical factors' and in the DSM-IV under the category of 'Mood disorders'. Limitations within current classifications for postnatal mental disorder are acknowledged.[58] In particular, limitations with the classification of PND according to the three main gold standard diagnostic systems available are that they differ with respect to the time frame of onset of the depressive episode after delivery. In addition, PND is not considered as an independent category or entity, but defined as an episode of depression that must occur within a relevant time frame post delivery. Accordingly, within the ICD-10, RDC and DSM-IV the symptoms and criteria for postnatal onset of a major or minor depressive episode do not differ from those of non-postnatal mood episodes. The set of criteria for classification of major depression are applied and if these are met then the subject receives the relevant diagnostic code plus an additional code that specifies postnatal onset.

TABLE I *Gold standard diagnostic classification systems: criteria for major depressive disorder*

	RDC	DSM-IV	ICD-10
Clinical significance	Prominent and relatively persistent dysphoric mood or pervasive loss of interest or pleasure characterised by symptoms: depressed, sad, blue, hopeless low, down in the dumps, 'don't care any more', irritable	Symptoms cause clinically significant stress or impairment in social, occupational or other important areas of functioning	Some or considerable difficulty in continuing with ordinary work, social or domestic activities, but will probably not cease to function completely, dependent on severity. Considerable distress and agitation, and unlikely to continue with social, work or domestic activities, except to a very limited extent, in severe episode
Duration of symptoms	Dysphoric mood present for at least 1 week. Probable episode if symptoms 1–2 weeks. Definite episode if symptoms > 2weeks	Symptoms present most of the day, nearly every day for at least 2 weeks	At least 2 weeks for all three grades of severity. If symptoms are severe and of very rapid onset diagnosis may be justified after 2 weeks
Criteria and severity	One or more distinct periods with dysphoric mood or pervasive loss of interest/pleasure. At least five of the following symptoms required as part of the episode for definite and four required for probable: (1) poor/increased appetite and/or weight loss/gain; (2) sleep difficulty/sleeping too much; (3) loss of energy, fatigability or tiredness; (4) psychomotor agitation or retardation; (5) loss of interest/ pleasure in usual activities, social contact or sex; (6) feeling of self-reproach or excessive/inappropriate guilt; (7) complaints/evidence of diminished ability to think or concentrate; (8) recurrent thoughts of death/suicide. Sought or was referred for help from someone during the dysphoric period, took medication, or had an impairment with family, at home, at school, at work or socially. Presents with none of the following, which suggest schizophrenia: (1) delusions of being controlled/influenced, or thought broadcasting, withdrawal, insertion; (2) non-affective hallucinations of any type throughout the day for several days; (3) auditory hallucinations; (4) delusions or hallucinations without any prominent depressive symptoms for > 1 month; (5) preoccupation with a delusion/hallucination to relative exclusion of other concerns; (6) definite instances of formal thought disorder, accompanied by blunted or inappropriate affect	Five or more of the following symptoms; at least one symptom is either depressed mood or loss of interest or pleasure: (a) depressed mood (e.g. feels sad or empty, or appears tearful; (b) marked loss of interest/pleasure in activities; (c) significant weight loss or gain, or decrease or increase in appetite; (d) insomnia or hypersomnia; (e) psychomotor agitation or retardation; (f) fatigue or loss of energy; (g) feelings of worthlessness or excessive or inappropriate guilt; (h) diminished ability to think or concentrate or indecisiveness; (i) recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or suicide attempt or a specific plan. For minor depressive disorder, at least two (but less than five) of the symptoms described above; at least one of the symptoms is either (a) or (b)	Typical symptoms of a depressive episode: (a) depressed mood; (b) loss of interest and enjoyment; (c) reduced energy leading to increased fatigability and diminished activity. Common symptoms of a depressive episode: (a) reduced concentration and attention; (b) reduced self-esteem and self-confidence; (c) ideas of guilt and unworthiness; (d) bleak and pessimistic views of the future; (e) ideas or acts of self-harm or suicide; (f) disturbed sleep; (g) diminished appetite. Typical somatic symptoms: (a) loss of interest/pleasure in activities that are normally enjoyable; (b) lack of emotional reactivity to normally pleasurable surroundings and events; (c) waking in the morning 2 hours or more before the usual time; (d) depression worse in the morning; (e) objective evidence of definite psychomotor retardation or agitation; (f) marked loss of appetite; (g) weight loss; (h) marked loss of libido. Criteria for severity of depressive episode: mild depressive episode: at least two of the typical symptoms, plus at least two of the other common symptoms present – episode diagnosed 'with somatic syndrome' if four or more somatic symptoms are present; moderate depressive episode: at least two of the typical symptoms, plus at least three (preferably four) of the other common symptoms present – episode diagnosed 'with somatic syndrome' if four or more somatic symptoms are present; severe depressive episode: all three of the typical symptoms should be present, plus at least four of the other common symptoms, some of which should be severe in intensity

DSM, Diagnostic and Statistical Manual of Mental Disorders; ICD, International Classification of Diseases, Injuries and Causes of Death; RDC, Research Diagnostic Criteria.

The DSM-IV does describe specific features of postnatal mood disorder in its accompanying text. The common symptoms associated with postnatal onset are described as mood lability, fluctuation in mood, guilt due to dissonance between the mother's mood and society's expectation of happiness, and disinterest or preoccupation with infant well-being. However, these descriptions do not form specific inclusion and exclusion criteria. Classification for diagnosis of PND is specified by the DSM-IV if onset of the depressive episode is within 4 weeks of the birth. In contrast, the ICD-10 defines PND as mental and behavioural disorders associated with the puerperium that commence within 6 weeks of delivery, and only if they cannot be classified elsewhere. The RDC, however, has never developed a specifier for postnatal major depressive disorder (Professor Jean Endicott, Columbia University Medical Centre, 2008, personal communication) and therefore does not specify criteria or a time frame for diagnosis of PND. Researchers often select their own time frame based on the DSM-IV or ICD-10 specifiers or consider use of an expert consensus opinion as to the time period that defines postnatal onset when applying the RDC.

With the future development of new classifications (the DSM-V is under development)[59] revisions suggested for inclusion in the new ICD classification system by a panel of international experts at the 1999 classification workshop in Sweden[60] were the introduction of a specifier for onset within 3 months postnatally that would cover all diagnoses of mood disorder, psychosis and adjustment disorder; omission of ICD-10 code F53 (mental and behavioural disorders associated with the puerperium, not elsewhere classified); the introduction of a further psychotic diagnostic category; and the introduction of a defined diagnostic category in the mood disorders section for subsyndromal or minor depression, also permitting the postnatal specifier. Nevertheless, the suggested revisions did not go so far as to suggest more specific inclusion and exclusion criteria for the classification of PND based on the specific descriptive symptoms for postnatal mood disorder observed in the DSM-IV text. The absence of a specific postnatal classification within any of the current diagnostic manuals is thought to reflect the underlying uncertainty in the entity of PND as a distinct diagnosis.[61]

Standardised clinical interview schedules

A range of psychiatric diagnostic interview schedules are available, which may be used to establish a diagnosis of depression or other psychiatric disorder based on the gold standard classification systems described previously. The need for clinical interview schedules arose because of the unreliability of psychiatric clinician diagnoses, for example during trials of the DSM-III, inter-rater reliability for major disorders ranged from kappa values of 0.28 to 0.92.[52] The aim of clinical psychiatric interviews is to distinguish using a standardised method between significant symptoms and the ordinary concerns and worries of daily life by setting requirements for clinical significance and distinguishing psychiatric symptoms from symptoms caused by drugs, alcohol and physical illness. Interview schedules may be structured or semistructured and provide a standardised method to increase the confidence in the diagnostic process and reliability of psychiatric diagnoses compared with open or unstructured interviews. Interview schedules provide greater inter-rater agreement between researchers and diagnosticians. Two types of interview schedule have been developed. The first type of interview schedule gives structure to the questions; however, the interviewer must make clinical judgements as to whether there is a need to probe with further questions based on whether the answers fulfil diagnostic criteria and therefore this type of interview schedule may be more suitable for administration by trained clinicians and clinically trained researchers. In the second type of interview schedule the questions are fully structured and interviewers are required to follow a fully specified route of questions; scoring is based on the subject's response and clinical judgement is not required, therefore these interviews are more suitable for trained lay interviewer administration. Standardised interview schedules are usually utilised within the research context rather than within the clinical context. Asking explicit questions about the symptoms based on stringent criteria ensures that a systematic and reliable diagnosis is established, and a replicable method ensures that any comparisons subsequently made across various research studies may be considered meaningful.[62–64] A number of interview schedules have been developed and brief summaries of six are given below.

Schedule for Affective Disorders and Schizophrenia

The Schedule for Affective Disorders and Schizophrenia (SADS) was designed for use alongside the RDC to formulate a diagnosis of current illness based on the defined inclusion and exclusion criteria for each RDC diagnostic category.[65] There are two other versions of the SADS interview: the lifetime version, which covers past episodes of mental disorder (SADS-L), and the version for measuring change (SADS-C). The questions are open-ended and do not require the respondent to restrict their answers to a yes/ no format only; follow-up questions are provided if the initial answer does not elicit enough information. Each symptom is assigned a rating between 0 and 7, reflecting the severity, intensity and pervasiveness: 0/1 represents no information/ not at all and 7 represents very extreme/constant. SADS comprises 26 items to determine a diagnosis of major or minor depression. Several studies have modified SADS for use with pregnant and postnatal women to take account of somatic symptoms that may result as a normal part of the postnatal experience, for example disturbed sleep due to the baby.

Structured Clinical Interview for DSM

The Structured Clinical Interview for DSM (SCID) is a semistructured clinical interview that is designed to formulate a diagnosis for mental disorders; the current clinical version is administered according to the DSM-IV criteria.[66] There are several versions of the SCID dependent on the population under assessment, for example psychiatric inpatients, outpatients and non-clinical populations. It is recommended that the SCID should be conducted by trained clinicians or experienced researchers as clinical judgement is required. The SCID comprises six self-contained modules; the clinical version takes approximately 45–90 minutes to complete and may be administered by paper and pencil method or alternatively software for administration and scoring is available.

Present State Examination

The Present State Examination (PSE) is a semistructured diagnostic interview that classifies cases of mental health problems according to the PSE-Index of Definition-Catego system, which is based on the ICD classification system.[67] The interview determines whether psychiatric symptoms have been present within the previous month and determines cases from non-cases according to the Catego index, which specifies the degree of certainty with which a subject may be considered a case. There are eight distinct levels within the index, each of which implies greater confidence in the classification as a psychiatric case; level five is considered the threshold that divides cases and non-cases. The Psychiatric Assessment Schedule (PAS) is an adaptation of the PSE that allows a subject to be classified according to the RDC. Computer software is available for the Catego index to fully automate the scoring of PSE data.

Standardised Psychiatric Interview

The Standardised Psychiatric Interview (SPI), also referred to as the Clinical Interview Schedule (CIS), is a semistructured clinical interview.[68] Its was designed for use in community surveys and is much briefer than other standardised psychiatric interviews. The SPI has also been modified with additional items relating to weight loss and appetite changes to allow classification according to the RDC. The SPI questions are designed to elicit the presence or absence of 10 psychiatric symptoms and the presence of an additional 12 manifest abnormalities of mental state, which are rated by the interviewer. Each psychiatric symptom is scored on a 5-point scale of severity; the scores for each of the 10 symptoms are then added to twice the scores of the 12 abnormalities of mental state to formulate a total score. The CIS has been revised (CIS-R)[69] and is a computerised version of the interview schedule that establishes the nature and severity of neurotic symptoms experienced over the previous 7 days and identifies the presence of neurosis and establishes a picture of overall health, appetite and physical activity. Each section scores a particular type of neurotic symptom (ranging in severity from 0 to 4), including, for example, somatic symptoms, fatigue, sleep problems, panic, depression, anxiety, compulsions and phobia. Symptoms with scores of 2 or more are considered clinically significant.

Diagnostic Interview Schedule

The DIS was developed as part of the 1978 Epidemiological Catchment Area programme. It was designed as a comprehensive, diagnostic interview for use in large-scale, multicentre epidemiological surveys and was developed because of the need to conduct surveys that would provide information regarding the prevalence

and incidence of specific psychiatric disorders in the USA.[70] Classification of mental disorders is according to the DSM criteria. The questions cover all of the symptoms necessary to make a diagnosis according to DSM criteria and ascertain lifetime history of symptoms in addition to the most recently experienced symptoms. The interview is fully structured and is suitable for trained lay interviewer administration; clinical judgements are not required and answers to the interview are precoded so that the interview data may be entered into computer software.

Composite International Diagnostic Interview

The CIDI is a comprehensive, fully structured diagnostic interview schedule for the assessment of mental disorders.[71] Lifetime and current diagnoses of disorders may be classified according to either the ICD or the DSM. It was designed for use in large-scale epidemiological surveys and is suitable for administration by trained lay interviewers via the paper and pencil method or computer (CIDI-auto). The questions are fully specified within the interview, with defined routes that the interviewers are required to follow, and do not require clinical judgements to be made by the interviewer. The responses are formatted as yes/no answers; positive responses to symptom questions are followed by questions from a probe flow chart that determine whether a symptom is a clinically significant psychiatric symptom or whether it is due to medication, drugs, alcohol or physical illness.

Mini-International Neuropsychiatric Interview

The Mini-International Neuropsychiatric Interview (MINI) is a short structured diagnostic interview for use with DSM-IV and ICD-10.[72] MINI contains 120 questions and covers 17 DSM-III-R axis I psychiatric disorders. It was developed jointly by psychiatrists and clinicians in the USA and Europe with the specific aim of reducing the administration and scoring time. The MINI has good correlation with other interview schedules, for example the kappa values for most psychiatric diagnoses with SCID-I and CIDI were 0.70 or above.[73,74]

Summary

We have provided an overview of the different standardised diagnostic interview schedules that have been developed to be conducted according to internationally recognised classification

systems. For the purposes of this review we define these approaches as providing a 'gold standard' in the diagnosis of PND. In addition to the use of diagnostic interview schedules there are other approaches to identify PND and these are described below. Of these approaches we have made a distinction between clinician and self-complete identification strategies.

Clinician-rated identification strategies

Clinician-rated scales are measures of depression used to standardise clinical judgements and provide ratings of duration and severity of symptoms. The measures are designed for use during a clinical consultation and are not suitable for large-scale population-based screening. Several clinician-rated scales are available to assess depression and monitor treatment response.

Raskin Depression Rating Scale

The Raskin Depression Rating Scale (RDRS; also known as the Raskin Three-Area Severity of Depression Scale) is a brief, clinician-rated scale suitable for assessing both baseline levels of depression and change in depression severity over time.[75] The scale takes 10–15 minutes to administer and requires the clinician to rate the patient's verbal report of symptoms in three areas: depressive symptoms (feeling low or downhearted, feeling worthless or helpless, loss of interest), depressed behaviour (looks sad, cries easily, psychomotor retardation, lacking energy) and secondary symptoms of depression that are primarily somatic (insomnia/hypersomnia, change in appetite, cognitive problems, thoughts/attempts of suicide). Items are rated on a 1–5 scale where 1 represents the response 'not at all' and 5 represents the response 'very much'.

Montgomery–Asberg Depression Rating Scale

The Montgomery–Asberg Depression Rating Scale (MADRS) was developed as an observer rating scale and is composed of 10 items. The items are concerned mostly with the psychological symptoms of depression and a global rating of degree of disturbance and social functioning is also included.[76] Each item is graded in severity from 0 to 6 and the total score ranges from 0 to 60. Scores between 7 and 18 indicate mild depression,

although a cut-off level of greater than 11 has been used.[33]

Hamilton Depression Rating Scale

The Hamilton Depression Rating Scale (HDRS; also referred to as the Hamilton Rating Scale for Depression – HRSD) was originally developed to assess the severity of depression among patients who had been diagnosed as depressed and was intended as a means of qualifying expert clinical opinion.[77] The original HDRS comprises 17 items on depressive symptoms, eight of which are concerned with somatic complaints; however, subsequent versions contain up to 31 items. Responses are rated on either a 3- or a 5-point scale with a score range between 0 and 50. The total score should be obtained from the sum of two independent ratings; if only one rater is used the score should be doubled for comparability. However, in practice, simple summative scores from one rater are widely used (without double scoring). A cut-off of 15 and above indicates major depression.

Self-report identification strategies

A wide range of self-report instruments are available for the identification of PND. These include generic depression strategies and postnatal-specific instruments. Generic depression (and sometimes anxiety) instruments are those designed and validated for the identification of depression in non-postnatal populations. Measures have also been developed specifically for use in postnatal populations. Both generic depression measures and postnatal-specific measures assess self-reported depressive symptoms and subjects rate their symptoms in terms of their frequency and severity.

Postnatal depression-specific strategies

Six postnatal-specific measures have been identified. Two of these measures [Pregnancy Risk Questionnaire (PRQ) and the Predictive Index (PI)] were developed for use prenatally to identify those women with depression during pregnancy and to identify those women at risk of development of significant depression in the postnatal period. A summary of each postnatal-specific measure is given in the following sections and in *Table 2*.

Edinburgh Postnatal Depression Scale
The EPDS was specifically developed to assist health professionals in the identification of depressive symptoms in community samples of postnatal mothers.[16] It is currently the most widely utilised self-report measure for the identification of PND.[30] During the developmental stage 21 items were selected for inclusion from several existing depression rating scales. These items were piloted on 100 women attending local health centres and as a result 13 items were selected as those most likely to identify PND; these included seven newly constructed items and six adapted items from the HADS and the Irritability, Depression and Anxiety (IDA) Scales.[78] The first 13-item version of the EPDS[79] was validated in a study conducted with 63 puerperal women using the SPI according to the RDC criteria diagnoses, with the interviewer blind to the EPDS scores. Examples of items included 'I have felt sad and miserable'

Table 2 Characteristics of post-natal depression specific strategies

Instrument	No. items	Score range	Time frame
Edinburgh Post-natal Depression Scale (EPDS)	10	0 to 30	Past 7 days
Post-natal Depression Screening Scale (PDSS)	35; 7 domains with 5 items	0 to 175	Past 14 days
Bromley Post-natal Depression Scale (BPDS)	10; plus chart to indicate when PND began, how long it lasted, and when it was worst	Unclear	Pregnancy and up to 1 year postnatally for each pregnancy
The Pitt Depression Scale (PDS)	24	0 to 48	Past few days
Pregnancy Risk Questionnaire (PRQ)	21; 18 ante-natal and 3 early post-natal components	18 to 90	Recently
Predictive Index (PI)	17	Unclear	Recently

and 'I have been so unhappy I have had difficulty sleeping'. Respondents were asked to describe the way they had been feeling for the past 7 days for each item on a 4-point scale. Although the 13 items distinguished between depressed and non-depressed cases, factor analysis found that three items formed a separate 'non-depressed' factor and therefore these were omitted from the final 10-item version.

The current 10-item version is scored on a 4-point scale (0–3) with a total score ranging from 0–30. Items written in the past tense include statements relating to maternal feelings in the past 7 days and refer to depressed mood, anhedonia, guilt, anxiety and suicidal ideation. The 10-item EPDS was validated in a study conducted with 84 women previously identified by health professionals as potentially depressed at 6 weeks postnatally. At 3 months postnatally a cut-off score of 12/13 correctly identified all 21 women with an RDC diagnosis of major depressive disorder with sensitivity, specificity and a positive predictive value of 86%, 78% and 73%, respectively, and therefore the cut-off for major PND was recommended as 12/13. The advantage of the EPDS, in contrast to generic depression measures, is that it does not include common somatic symptoms such as insomnia and appetite changes, which may occur normally within the postnatal period. The EPDS is usually administered by the paper and pencil method although computerised versions are now available; it is brief, taking approximately 5 minutes to complete, and easy to administer, interpret and score.

Postnatal Depression Screening Scale
The PDSS is a 35-item self-report measure created specifically for new mothers that can be administered in 5–10 minutes; it is brief and easy to understand and interpret.[17] The conceptual basis of the PDSS was through a series of qualitative studies of PND. The PDSS consists of seven dimensions, each of which contains five items. The dimensions include sleeping/eating disturbances, anxiety/insecurity, emotional lability, cognitive impairment, loss of self, guilt/shame, and thoughts of self-harm. Each item describes the type of feelings a woman may experience after the birth of a child. Respondents are asked to indicate their agreement or disagreement on a 5-point scale ranging from 'strongly agree' to 'strongly disagree' regarding how they have felt in the last 2 weeks. The identification strategy measure yields an overall severity score falling into one of three ranges: normal adjustment, significant symptoms

of PND and positive screen for major PND. A study of 150 mothers at 12 weeks post delivery who completed the PDSS, BDI and EPDS found that the PDSS was strongly correlated with the BDI ($r = 0.81$) and the EPDS ($r = 0.79$).

Bromley Postnatal Depression Scale
The Bromley Postnatal Depression Scale (BPDS) was specifically developed as a method to identify the presence of both current and previous episodes of PND.[80] In particular, the main purpose was to devise a method that would be suitable for diagnosing PND retrospectively, following previous pregnancies to build a longitudinal picture of depressive disorder within the postnatal period. The conceptual basis for the BPDS was formed following interviews with women who had experienced PND some years previously, but who were currently attending a psychiatric outpatient clinic for other reasons. The women were not able to reliably recall the presence or absence of individual symptoms but were able to describe a global impression of low mood, feeling unwell and impaired functioning during the postnatal period. The authors constructed a vignette based on a description of depression following childbirth by Pitt;[81] this forms the basis of the questionnaire. Respondents are requested to read the vignette and then answer seven questions that ask them to recall past experiences of PND; for example, questions with a yes/no response include 'Did you suffer from postnatal depression after the birth of the first baby?', 'Did you take any tablets or medications for depression or nerves?' and 'Were you admitted to a psychiatric hospital or ward in the first year after the birth of this baby?' Three further questions include the recall of information regarding depressed feelings, taking medication and admission to psychiatric care during pregnancy. The authors validated the BPDS against the Dunedin scale, a questionnaire that examines feelings within the first year of childbirth, which was validated against the DSM-III criteria for major depression. Assuming that a positive response to the Dunedin scale was equal to a DSM-III diagnosis, then the sensitivity and specificity of the BPDS was 62% and 94%, respectively, for sense.

The Pitt Depression Scale
The Pitt Depression Scale (PDS) represents a 24-item questionnaire based on clinical experience and measures maternal anxiety and depression before and after childbirth.[81] The items are listed as questions, for example 'Do you worry a lot about the baby?' and 'Are you as happy as you ought to be?' The respondent indicates whether each

symptom was present 'today, or over the past few days' and responds 'yes', 'no' or 'don't know'. The total scores range from 0 to 48. The PDS correlates highly with the EPDS; however, it has not been validated and remains infrequently used.

Pregnancy Risk Questionnaire
The PRQ is a self-report measure that was developed from a review of salient risk factors associated with PND and on the basis of their face and construct validity. The questionnaire contains 18 antenatal components and three early postnatal components. Items are listed as questions and women are asked to circle their responses to the questions on a 5- or sometimes 6-point Likert scale. Reponses correspond to the categories 'not at all', 'somewhat' and 'very much'. For example, the first question is, 'Overall, has this pregnancy been a positive experience for you?'

Predictive Index
The PI is a 40-item self-report questionnaire that has been developed from key predictive factors identified from two large British prospective epidemiological surveys.[11,82] The questionnaire assesses six domains: (1) the emotional experience of the pregnancy (whether it was viewed positively and whether anxiety or depression was experienced); (2) the physical experience of the pregnancy (whether medical help was sought for conception, or any health problems were encountered); (3) psychiatric history (previous depression history and past experience of PND); (4) maternal bereavement, before 11 years of age; (5) the quality of social relationships (current quality of the woman's relationship with her mother and partner, length of relationship with partner and access to other confidants); and (6) social factors (satisfaction with living area, educational achievement and feelings about giving up work). Responses to the questionnaire are categorical and different scoring criteria are available for primiparous and multiparous women, and for women who have experienced a previous episode of PND.

Generic depression strategies

A number of generic depression measures are known to exist.[83] *Table 3* describes the characteristics of some of the widely used generic depression (and sometimes anxiety) instruments.

Case-finding questions

In some areas of research there has been a shift away from using self-report strategies to using case-finding questions (e.g. for depression[89]). The PRIME-MD[90] was designed to identify depression in primary care and classifies patients according to the DSM-IV criteria. The PRIME-MD contains two initial questions about depressed mood and anhedonia that may be asked during consultation: (1) 'During the past month, have you often been bothered by feeling down, depressed or hopeless?' and (2) 'During the past month, have you often been bothered by little interest or pleasure in doing things?' A positive response to one of these questions prompts the clinician to ask the patient to complete the patient questionnaire of the PRIME-MD, a one-page self-report questionnaire that assesses five dimensions of psychiatric disorders, including mood disorders, and

TABLE 3 *Characteristics of generic depression strategies*

Instrument	Items[a]	Scope	Time frame	Score range	Administration time
BDI[18]	21, 13, 7	Depression-specific	Today	0–63	5 minutes
GHQ[84]	30, 28, 12	Global psychiatric illness (including depression and anxiety)	Past few weeks	0–28	2–10 minutes
HADS[85]	14	Depression and anxiety	Past week	0–21	≤ 2 minutes
HSCL[86]	25, 13	Global with depression-specific category	Past week	25–100	2–5 minutes
SCL-90-R[87]	90	Global with depression-specific category	Past week	0–360	12–15 minutes
Zung's SDS[88]	20	Depression-specific	Recently	25–100	2–5 minutes

BDI, Beck Depression Inventory; GHQ, General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; HSCL, Hopkins Symptom Checklist; SCL-90-R, Symptom Checklist-90-Revised; Zung's SDS, Zung's Self-rating Depression Scale.
a   Different numbers refer to different versions.

19

comprises 26 items regarding symptoms and one item regarding general health.

An alternative way of using these questions is to simply ask patients to give 'yes' or 'no' responses and use this as the identification strategy rather than as a prompt for further investigation. Whooley *et al.*[38] have demonstrated the use of the two brief questions, developed from the 2-item PHQ-9, compared with usual measures for identifying depression in a group of male participants attending a Veterans Affairs Medical Center.

Summary

Measures used in the identification of PND are numerous. Diagnostic measures such as the DSM, ICD and RDC are considered the gold standard reference case for classification of mood disorders within the postnatal period. However, the current classification of PND as a mood disorder within the postnatal period reflects underlying uncertainty regarding the entity of PND as a distinct diagnosis. In addition, disparities in diagnosis may arise because of differences in the time frame specified for onset and in the criteria for a major depressive disorder between the classification systems, for example more symptoms must be established using the RDC than with the DSM. There are numerous generic and postnatal-specific measures that may be used to identify possible cases of PND.

Reflection on current policy and practice within the UK

As part of the NICE guidance[30] issued on antenatal and postnatal mental health a survey of primary care trusts (PCTs) in England and Wales was undertaken. The guideline development group sent a brief questionnaire to all PCTs in England and Wales with the aim of gaining an understanding of current service provision within primary care. As part of this survey, 64% of respondents included free-text comments; within these comments, 40% reported using the EPDS as an assessment tool (93% of those mentioning such tools). Despite the widespread use of the EPDS, the NICE guidance recommends using the case-finding questions developed by Whooley and colleagues with the additional help question ('Is this something you feel you need or want help with') if women respond yes to either of the two Whooley questions.

Chapter 5
Validity of methods to identify postnatal depression: systematic review 1

Key concepts in diagnostic accuracy studies

In the previous chapter a number of methods to identify PND were outlined. In this chapter we summarise the available evidence regarding the validity of these methods. Diagnostic accuracy studies aim to measure the amount of agreement between index test results and the outcome of the reference standard. When we focus on PND, for the purposes of this review, the reference standard was a standardised diagnostic interview conducted according to internationally recognised criteria. Hence, the identification strategy would be administered to a series of women in the pre- or postnatal period, and the presence or absence of PND would be determined by the outcome of the diagnostic interview.

In general, when any test is used there are four possible outcomes:

- when a person has the condition the test may be positive (true positive)
- when a person has the condition the test may be negative (false negative)
- when a person does not have the condition the test may be negative (true negative)
- when a person does not have the condition the test may be positive (false positive).

The results of a diagnostic accuracy study are often summarised in a 2×2 table, as shown in *Table 4*.

Several measures of a test's performance can be calculated from this summary information.

Two frequently used measures are sensitivity and specificity. Sensitivity of the test is the probability of a positive test result using the index test given the individual has the target condition. For example, the sensitivity of the EPDS is the proportion of women who score above a predefined threshold who have PND as classified using the diagnostic interview. Specificity of the test is the probability of a negative test result using the index test given the individual does not have the target condition. For example, the specificity of the EPDS is the proportion of women who score below a predefined threshold who do not have PND as classified using the diagnostic interview.

Sensitivity and specificity of an identification strategy vary as a function of the cut point used. A cut point is used to indicate which individuals are likely to have the target condition (e.g. those scoring at or above the cut point) and which individuals are unlikely to have the target condition (e.g. those scoring below the cut point). For example, if a cut point of 13 is used with the EPDS then women scoring 13 or above would be grouped as having PND, whereas those women scoring 12 or below would be grouped as not having PND. The EPDS is scored on a scale from 0 to 30 and any value in this range could be used as a cut point. Sensitivity and specificity are dependent upon one another – if one value decreases the other value increases. Hence, increasing the cut point increases or decreases the sensitivity and specificity of the test. Youden's index is one way of attempting to summarise test performance into a single numeric value to aid decision-making regarding cut points and was the method chosen to define the optimum cut point in this chapter.[91]

TABLE 4 Summary 2×2 table

Index test	Reference test +	–
+	True positive	False positive
–	False negative	True negative

Methods

Inclusion criteria

Two reviewers screened titles and abstracts to identify potentially eligible studies. Any disagreements were resolved by consensus or deferred to a third party if necessary. Full papers for potentially eligible studies were obtained and assessed for inclusion independently by two reviewers. Articles were eligible for inclusion if they fulfilled the following criteria:

- population: women in the prenatal or postnatal period (up to 1 year)
- setting: all settings
- identification test: any standardised depression screening/case-finding instrument or standardised clinical assessment tool
- reference test/gold standard: a standardised diagnostic interview conducted according to internationally recognised criteria, such as the ICD system, versions of the DSM or the RDC and specific primary care versions of this diagnostic system (e.g. the PRIME-MD)
- design: cross-sectional, case–control (case–referent), cohort studies and randomised controlled trials (RCTs), in which instruments were used at inception as a method of recruitment.

Categorisation of studies

Studies were examined separately by the type of identification strategy (prenatal or postnatal) and disease classification (major depression only, major or minor depression, any psychiatric disorder, or other) used. Within the articles retrieved a number of different systems were used to define cut points. To maintain consistency and permit pooling we classified a cut point of x as being a woman scoring x or above on the identification strategy used. Some studies reported cut points differently. For example, when using the EPDS, cut points were reported as 13 or 12.5 or 12/13 or > 12. In all of these examples we would have classified them as using a cut point of 13. This analogy was extended to other identification strategies and cut points.

In many cases multiple data points were presented for the studies evaluated. Multiple data points arose for many reasons, singularly or in combination: more than one identification strategy was used; more than one cut point was presented; the identification strategy was repeatedly administered; two versions of the identification strategy were used (e.g. the original strategy and a shorter version, or English and Punjabi versions of the identification strategy); different classifications of depression were recorded; or more than one reference standard was used. Studies were pooled at individual cut points to attempt to overcome the fact that multiple data points were presented.

Quality assessment

The quality of studies was assessed according to accepted criteria, Quality Assessment of Diagnostic Accuracy Studies (QUADAS). QUADAS is a structured checklist comprising of 14 items which are recorded as 'yes', 'no' or 'unclear'. The items provide a standardised approach to quality assessment and cover patient spectrum, choice of reference standard, disease progression bias, verification bias, review bias, clinical review bias, test execution, study withdrawals and indeterminate results. Two reviewers independently rated the quality of studies using QUADAS. The quality of non English language (e.g. German) studies was only assessed with QUADAS by a single reviewer.

QUADAS items were rated as 'yes', 'no' and 'unclear' in accordance with the user's guidance.[92] Any disagreement between reviewers was resolved by discussion or consensus by a third party. QUADAS items for which articles had been preselected or which were not applicable were excluded. Hence, item 12 was excluded from QUADAS as scoring of the index test should be fully automated. In relation to question 4, a 2-week period between the reference standard and index test was regarded, a priori, as short enough to be reasonably sure that the target condition did not change between the two tests. Question 13 was altered slightly to refer to missing items/unclear responses on the identification strategy rather than uninterpretable/intermediate test results.

We assessed agreement between the two reviewers. Kappa statistics were calculated for each question to assess inter-rater reliability. The following guidelines were used to interpret the strength of agreement: < 0.2 = poor, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = good, and 0.81–1.00 = very good.[93]

Data extraction

For all English language articles data were extracted independently by two reviewers. Non-English language papers were extracted by one reviewer, accompanied by a translator. Data extracted from non-English language studies were not assessed by a second reviewer.

Data synthesis

Studies that reported the results of applying the same identification strategy using the same cut point to diagnose the same type of disorder were pooled using a bivariate analysis. Pooled estimates of sensitivity, specificity, positive and negative likelihood ratios and diagnostic odds ratio (DOR), together with associated 95% CIs, were calculated. The model was fitted using a generalised linear mixed model approach to the bivariate meta-analysis of sensitivity and specificity.[94] This approach uses the exact binomial distribution to describe the within-study variability of sensitivity and specificity rather than the normal approximation, which was originally proposed.[95] Hence, it is preferable when there are low cell counts. The generalised linear mixed model approach that we used corresponded to the approach to fitting the hierarchical summary receiver operating characteristic (HSROC) model.[96]

The bivariate approach preserves the two-dimensional nature of the original data by analysing pairs of sensitivity and specificity data jointly, incorporating any correlation that might exist between these two measures using a random-effects approach. The bivariate approach fits a two-level model, with independent binomial distributions for the true positives and true negatives conditional on the sensitivity and specificity in each study, and a bivariate normal model for the logit transforms of sensitivity and specificity between studies.

Tests for heterogeneity were carried out for each outcome and are reported below. Between-study heterogeneity was assessed using the I^2 statistic of the pooled DOR. The I^2 statistic quantifies the degree of inconsistency in the studies' results by describing the percentage of total variation across studies due to heterogeneity rather than chance.[97] The I^2 statistic has advantages over other measures of heterogeneity (such as chi-squared test), including greater statistical power to detect clinical heterogeneity when fewer studies are available. The I^2 statistic ranges from 0% to 100% with a value of 0% indicating no heterogeneity; the further away from 0% the I^2 statistic is the more heterogeneous the set of studies are. To aid interpretation, tentative categorisations have been suggested: 25% = low heterogeneity, 50% = moderate heterogeneity, 75% = high heterogeneity.[98] These categorisations were adopted in this review.

If significant between-study heterogeneity was present we sought to explore the causes of this heterogeneity. First, summary ROC (sROC) curves were constructed using the bivariate model to produce a 95% confidence ellipse within ROC space (see Appendix 4). Summary ROC plots were visually inspected to identify those studies that lay outside the 95% confidence ellipse. Second, further analyses were conducted using D [log (DOR)]. A weighted multivariate linear meta-regression analysis was used, with weights proportional to the reciprocal of the variance of D representing the within-study variation (using restriction maximum likelihood estimation). Clinical variables that were considered, a priori, were time since birth and baseline prevalence of depression. The effect of quality features was also examined (individually as opposed to a summary quality score) by including method of verification and blinding of reference standard. If these items were important sources of heterogeneity then they would have been predictive in a meta-regression analysis and would have reduced the level of between-study heterogeneity in a meta-regression model. For dichotomous predictor variables, the meta-regression model produced a 'ratio of diagnostic odds ratios', in which deviation from 1 suggests a difference in the pooled estimates according to the predictor variable. Following recommendations, meta-regression was only undertaken if there were 10 or more studies included in the analysis.[99] All analyses were conducted using STATA version 9,[100] including the user-written STATA commands *metandi*[101] and *glamm*.[102]

Results

A total of 64 articles met the inclusion criteria and provided sufficient data to calculate full 2×2 cross-tabulations. Four articles either provided data from the same sample of women, but at different time points or used different instruments or provided additional data at different cut points. Hence, the 60 studies were reported in 64 articles (see *Tables 7* and *9*). Of the 60 studies, four were published in languages other than English: two in Spanish,[103,104] one in German[105] and one in Japanese.[106]

Characteristics of included studies
Classification

In total, 27 studies focused on women with major depression (DSM or equivalent) only, 39 studies focused on major or minor depression (DSM or equivalent), nine studies focused on any psychiatric disorder and three studies focused on other types of disorders (*Table 5*).

TABLE 5 *Type of depression by when the identification strategy was administered*

Type of depression	Prenatal	Pre- and postnatal	Postnatal	Unclear	Total
Major only	4	0	9	0	13
Major or minor	0	1	20	1	22
Major or minor and major only	2	0	12	0	14
Major or minor and other	1	1	1	0	3
Any psychiatric disorder	1	0	8	0	9
Other	0	0	3	0	3
Total	8	2	53	1	64

A single study was recorded as unclear regarding whether the identification strategy was administered pre- or postnatally.[107] In this particular study, 54 first-time mothers were asked to complete the Portuguese version of the EPDS and were interviewed using SADS (diagnosis according to RDC) antenatally at 6 months' gestation and at 12 months postnatally, with a subsample additionally interviewed at 3 months postnatally. The psychometric attributes (sensitivity, specificity, positive predictive value, negative predictive value) were presented across all of the time points for a variety of cut points. Sensitivity values ranged from 0.29 to 0.79 and specificity values ranged from 0.83 to 0.96. An optimal cut point, in terms of the trade-off between sensitivity and specificity, was 9, giving a sensitivity of 0.71 and a specificity of 0.89. As the identification strategy was administered pre- and postnatally and the results were combined across these periods this study could not be incorporated any further in the analysis.

Quality assessment
Kappa statistics were calculated for each quality assessment question to assess agreement between reviewers (*Table 6*). The overall proportion of agreement between the two reviewers for all QUADAS items combined was 0.85 (Kappa 0.69) indicating good agreement. The proportion of agreement between reviewers for each item ranged from 0.55 to 1.00 and was over 80% for eight of the items. The poorest agreement was associated with the items for selection criteria (item 2), indeterminate results (item 13) and withdrawals (item 14). Examination of cross-tabulated data revealed that disagreement was generally between 'yes' and 'unclear' responses or 'no' and 'unclear' responses, rather than between 'yes' and 'no' responses.

Prenatal results
This section focuses on the 10 retrieved studies (4236 women) that administered identification strategies in the prenatal period to identify women with depression and/or women at risk of developing depression in the postnatal period.

Quality assessment
Quality assessment was undertaken using QUADAS and the results are shown in *Figure 2*. There was variability in the results of the quality assessment. Studies did well in five out of the eight questions focusing on bias (questions 3, 5, 6, 7, and 14) and in all three questions relating to reporting quality (questions 8, 9 and 13). Over 70% of studies scored 'yes' in answer to these questions.

Several QUADAS items were poorly described in the diagnostic studies. This included both questions relating to variability: in 50% of studies the spectrum of participants was not representative and in 60% of studies it was unclear how participants were selected for the study. Question 10 had the lowest quality rating. It was unclear from 80% of the studies whether the index tests were interpreted without knowledge of the reference standard. Interestingly, the studies performed better on question 11, a related question; for this question it was unclear in 40% of studies whether the reference standard was interpreted without knowledge of the index test.

Characteristics of included studies
Studies were published between 1990 and 2006 and were undertaken in a variety of countries: two in Nigeria,[108,109] two in the UK,[19,110] and a single study each in France,[111] Australia,[112] the USA,[113] Tanzania,[114] Japan[115] and Malta.[116] The percentage of women with PND in these studies ranged from 5% to 25%, according to the reference standard

TABLE 6 *Agreement between reviewers when applying QUADAS*

Item	Proportion of agreement	Kappa statistic
Was the spectrum of patients representative of the patients who will receive the test in practice?	0.70	0.37
Were selection criteria clearly described?	0.55	0.27
Is the reference standard likely to classify the target condition correctly?	0.98	0.66
Is the period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	0.77	0.59
Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?	0.80	0.51
Did patients receive the same reference standard regardless of the index test result?	0.98	0.00
Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	1.00	—[a]
Was the execution of the index test described in sufficient detail to permit replication of the test?	0.93	−0.03
Was the execution of the reference standard described in sufficient detail to permit its replication?	0.90	0.64
Were the index test results interpreted without knowledge of the results of the reference standard?	0.92	0.77
Were the reference standard results interpreted without knowledge of the results of the index test?	0.95	0.90
Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	—[b]	—[b]
Were uninterpretable/intermediate test results reported?	0.73	0.12
Were withdrawals from the study explained?	0.75	0.32

a   Unable to calculate as too few rating categories.
b   Item 12 was excluded as scoring of the index test should be fully automated.
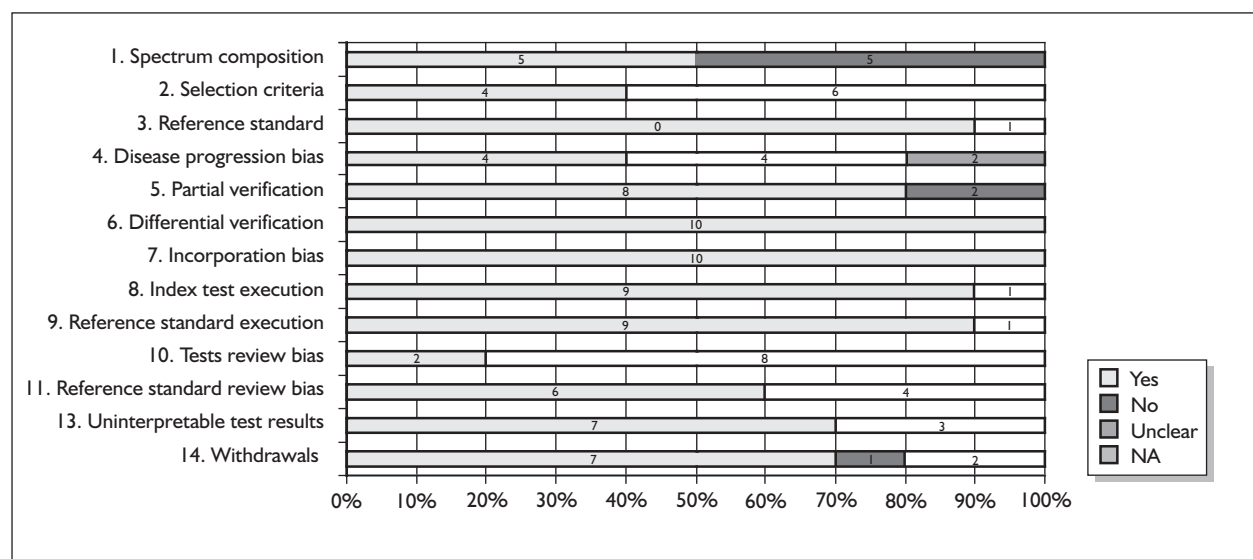
FIGURE 2 *Summary of quality assessment for prenatal studies.*

criteria. Eight of the studies[19,108,109,111,113–116] aimed to validate identification strategies for prenatal use by administering the identification strategy and reference standard during pregnancy. The two remaining studies[110,112] also administered the identification strategy prenatally, but administered the reference standard postnatally to measure the predictive strength of the identification strategy.

Eight different instruments were reported as being used prenatally across the diagnostic accuracy studies included in this part of the review. *Figure 3* displays the number of studies utilising the different instruments. It is clear that the EPDS was the most frequently used instrument. A number of the instruments have been translated into other languages and validated. For example, the EPDS has been translated from English into three other languages (*Table 7*). As previously stated there are multiple reference standards that can be used to establish whether an individual may be depressed or not and the type of depression that they have. In this review, five (50%) of the studies used DSM, two (20%) used ICD and two (20%) used RDC classifications. It was unclear in the final study which criteria were used; it was only reported that the CIDI interview schedule was used.

In the diagnosis of post-natal depression, as with other mental health conditions, cut points on the scores from paper and pencil based questionnaires are often chosen to distinguish between cases and non-cases or major and minor episodes. In practice, careful consideration needs to be given to decide what the most appropriate cut point to be used is, as the cut point chosen affects the accuracy of the test. Choosing a lower cut point to distinguish between post-natal depression cases and non-cases will lead to higher sensitivity but lower specificity values. A more sensitive test will result in fewer women with PND being unidentified, however, the lower specificity of the test will result in more women being wrongly identified as having PND.

In the original validation study of the EPDS a cut point of 13 was recommended as the most likely cut point to identify women suffering from a depressive illness of varying severity.[16] It was also recommended that a cut point of 10 should be used if the scale was considered for routine use. Table 8 summarises the cut points reported in the studies using the EPDS to identify women likely to have the target condition. The studies were summarised by the type of disorder the gold standard diagnostic interview was used to classify (i.e. major depression only, major or minor depression, any psychiatric disorder and other categories). A variety of cut points were reported across all studies using the EPDS, ranging from 0 to 26; however, for all classifications the most frequently reported cut point was 13.

Major depression (DSM or equivalent) only

Results for the diagnosis of major depression only were reported in six out of the 10 prenatal studies.[19,109–112,114] Two of these studies[110,112] were
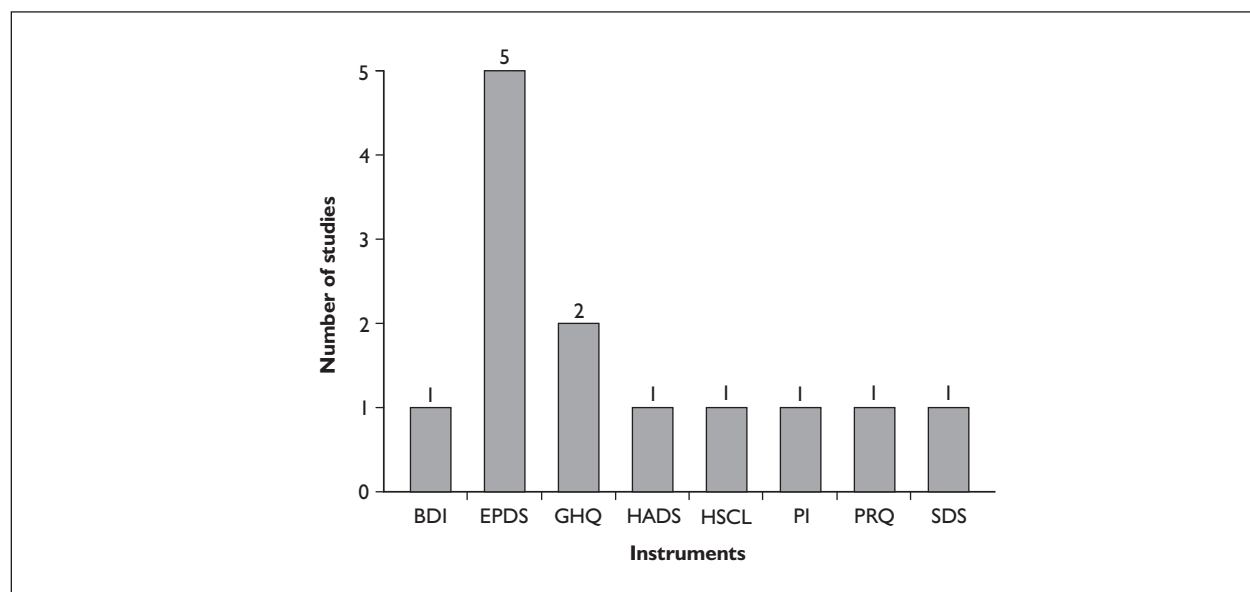

FIGURE 3 *Number of included studies using different instruments prenatally. BDI, Beck Depression Inventory; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; HSCL, Hopkins Symptom Checklist; PI, Predictive Index; PRQ, Pregnancy Risk Questionnaire; SDS, Zung's Self-rating Depression Scale.*

TABLE 7 Characteristics of included studies using identification strategies prenatally

Study, country, study type	Version	Study setting, validation sample	Identification strategy	Time of admin., setting if different	Interview type, gold standard	Type of depression	n (%) with depression	Sample size
Abiodun, 1994,[108] Nigeria, validation	English or Yoruba	Antenatal clinics at teaching hospital, all	GHQ-12 HADS-depression	Unclear	PSE, ICD-9	Other Major or minor	11 (4.6)	240
Adewuya et al., 2006,[109] Nigeria, validation	English or Yoruba	Antenatal clinics at teaching hospital, EPDS ≥ 12 + 10% EPDS < 6	EPDS	32–36 weeks' gestation	MINI, DSM-IV	Major only Major or minor	9 (10.5) 15 (17.4)	86
Adouard et al., 2005,[111] France, validation	French	Antenatal clinic at hospital specialising in high-risk pregnancies, all	EPDS	28–34 weeks' gestation	MINI, DSM-IV	Major only	15 (25.0)	60
Felice et al., 2006,[116] Malta, validation[a]	Maltese	Antenatal clinics at hospital, all	EPDS	19 weeks' gestation	CIS-R, ICD-10	Major or minor	32 (14.3)	223
Holcomb et al., 1996,[113] US, validation	English	Women's health centre, obstetric complications clinic, centre for diabetes and pregnancy, fetal testing unit of teaching hospital, all	BDI	32 weeks' gestation, unclear	NIMH-DISC-III, DSM-III-R	Any psychiatric disorder	12 (11.4)	105
Kaaya et al., 2002,[114] Tanzania, Validation	Kiswahli	Women included in RCT, first 50 HSCL-25 > 1.75 + random 50 HSCL-25 ≤ 1.75	HSCL-15, HSCL-25, HSCL-R	18 weeks' gestation	SCID, DSM-IV	Major only	11 (11.1)	99
Kitamura et al., 1994,[115] Japan, validation[a]	Japanese	Obstetric department of general hospital, all	SDS GHQ-30	First trimester 34 weeks' gestation First trimester 34 weeks' gestation	SADS, RDC	Major or minor Other	11 (9.9) 10 (9.8) 18 (16.7) 13 (13.3)	111 102 108 98

TABLE 7 Characteristics of included studies using identification strategies prenatally (continued)

Study, country, study type	Version	Study setting, validation sample	Identification strategy	Time of admin., setting if different	Interview type, gold standard	Type of depression	n (%) with depression	Sample size
Murray and Cox, 1990,[19] UK, validation	English	Antenatal clinic of large maternity hospital, all	EPDS	28 and 34 weeks' gestation	SPI, RDC	Major only	6 (6.0)	100
						Major or minor	14 (14.0)	100
Austin et al., 2005,[112] Australia, prediction	English	Midwife-based clinic at large obstetric hospital, ≥ 12 or reported felt so miserable or sad that interfered with ability to get things done or their relationships with family/friends + 7% random others	EPDS, PRQ – antenatal section	32 weeks' gestation, posted back, and postal questionnaire 2–4 months postnatally	Computerised CIDI, unclear	Major only	69 (5.3)	1296
Cooper et al., 1996,[110] UK, prediction	English	Antenatal clinics at teaching hospital, EPDS > 8	PI	32 weeks' gestation and postal questionnaire 6–8 weeks postnatally	SCID, DSM-III-R postnatal	Major only	293 (15.3)	1916

BDI, Beck Depression Inventory; CIDI, Composite International Diagnostic Interview; CIS, Clinical Interview Schedule; DSM, *Diagnostic and Statistical Manual of Mental Disorders*; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; HSCL, Hopkins Symptom Checklist; ICD, International Classification of Diseases; MINI, Mini-International Neuropsychiatric Interview; NIMH-DISC, National Institute of Mental Health Diagnostic Interview Schedule; PI, Predictive Index; PRQ, Pregnancy Risk Questionnaire; PSE, Present State Examination; RCT, randomised controlled trial; RDC, Research Diagnostic Criteria; SADS, Schedule for Affective Disorders and Schizophrenia; SCID, Structured Clinical Interview for DSM; SDS, Zung's Self-rating Depression Scale; SPI, Standardised Psychiatric Interview.
a Studies also validated identification strategies postnatally.

Table 8 *Summary of the type of depression by the cut points used for the EPDS*

Cut point	Major only	Major or minor	Any psychiatric disorder	Other
0	0	0	1	0
1	1	0	0	0
2	1	0	1	0
3	1	1	0	1
4	1	2	1	1
5	1	2	0	1
6	1	3	2	1
7	4	10	2	2
8	9	14	3	2
9	10	20	4	2
10	14	22	4	6
11	12	21	4	2
12	15	20	5	3
13	18	24	7	7
14	8	12	4	2
15	7	10	2	0
16	4	6	3	0
17	2	3	2	0
18	1	1	2	0
19	1	1	1	0
20	1	0	2	0
21	1	0	1	0
22	1	0	2	0
23	1	0	0	0
24	1	0	1	0
25	1	0	1	0
26	1	0	0	0

the prediction studies and were not combined with the results of the validation studies. Both studies administered the identification strategy at 32 weeks' gestation and posted women the identification strategy to complete in the postnatal period. The first study used the PI during pregnancy.[110] At 6–8 weeks postnatally women were posted the EPDS to complete and return. Women scoring 8 or above on the EPDS were subsequently visited at home and a diagnostic interview undertaken. The sensitivity and specificity were reported for a range of cut points (*Figure 4*). Sensitivity ranged from 0.05 (specificity 0.98) to 1 (specificity 0.06), with the optimal cut point at 23, in terms of the trade-off between sensitivity and specificity, giving a sensitivity of 0.59 and a specificity of 0.67.

The second study used the EPDS and the antenatal section of the PRQ.[112] Women were administered the EPDS and the PRQ during pregnancy and then reviewed by postal questionnaire at 2 and 4 months postnatally using the EPDS. Women scoring 12 or above on the EPDS or reporting that in the last 2 months there had been a period of 1 week or more when they felt so miserable or sad that it interfered with their ability to get things done or interfered with their relationships with friends/ family were given a diagnostic interview over the telephone. The authors reported the sensitivity and specificity at three cut points for each identification strategy (*Figure 5*). Sensitivity for the EPDS ranged from 0.22 (specificity 0.91) to 0.78 (specificity 0.49). Sensitivity for the PRQ ranged from 0.44 (specificity 0.92) to 0.81 (specificity 0.60).
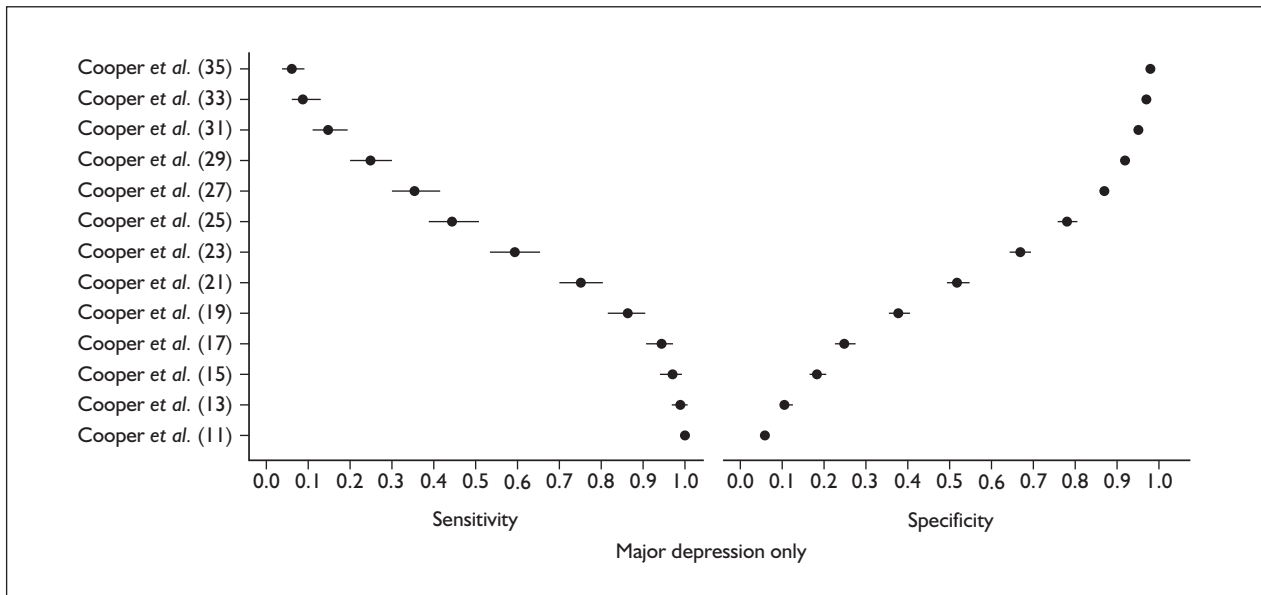
FIGURE 4 *Single study using the Predictive Index prenatally. Values in parentheses on the left-hand side of the graph represent the various cut points used to define major depression.*
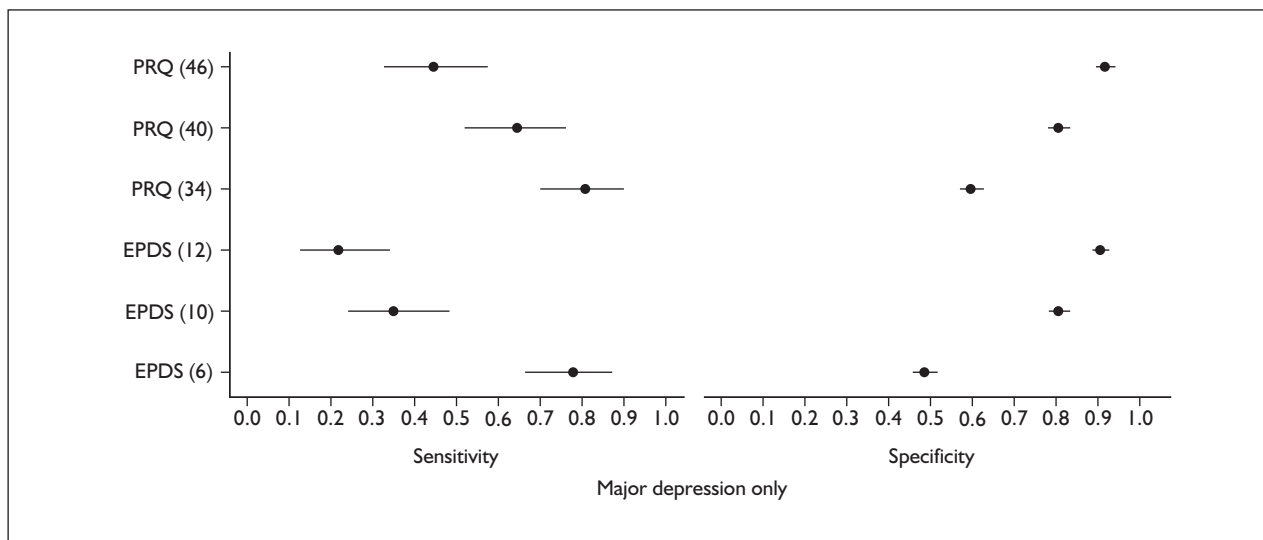

FIGURE 5 *Single prediction study using the PRQ and EPDS prenatally. Values in parentheses on the left-hand side of the graph represent the various cut points used to define major depression. EPDS, Edinburgh Postnatal Depression Scale; PRQ, Pregnancy Risk Questionnaire.*

Three of the remaining four validation studies used the EPDS[19,109,111] and the final study used the Hopkins Symptom Checklist (HSCL).[114] Unfortunately, there were insufficient data at each cut point for the EPDS and the HSCL to permit pooling. A summary of the sensitivity and specificity of the studies using the EPDS and the HSCL are summarised in *Figures 6* and *7* respectively. Sensitivity for the EPDS ranged from 0.56 (specificity 1.00) to 1.00 (specificity 0.79). Sensitivity for the HSCL was 0.89 with specificity ranging from 0.79 to 0.85.

Major or minor depression (DSM or equivalent)

Five out of the 10 prenatal validation studies focused on major or minor depression. Three studies used the EPDS,[19,109,116] one study used the HADS[108] and the final study used the Zung's SDS.[115] There were insufficient data for each instrument at the various cut points to permit pooling of diagnostic data. A summary of the sensitivities and specificities for the EPDS and Zung's SDS are shown in *Figures 8* and *9* respectively. Sensitivity for the EPDS ranged from 0.50 (specificity 0.99)
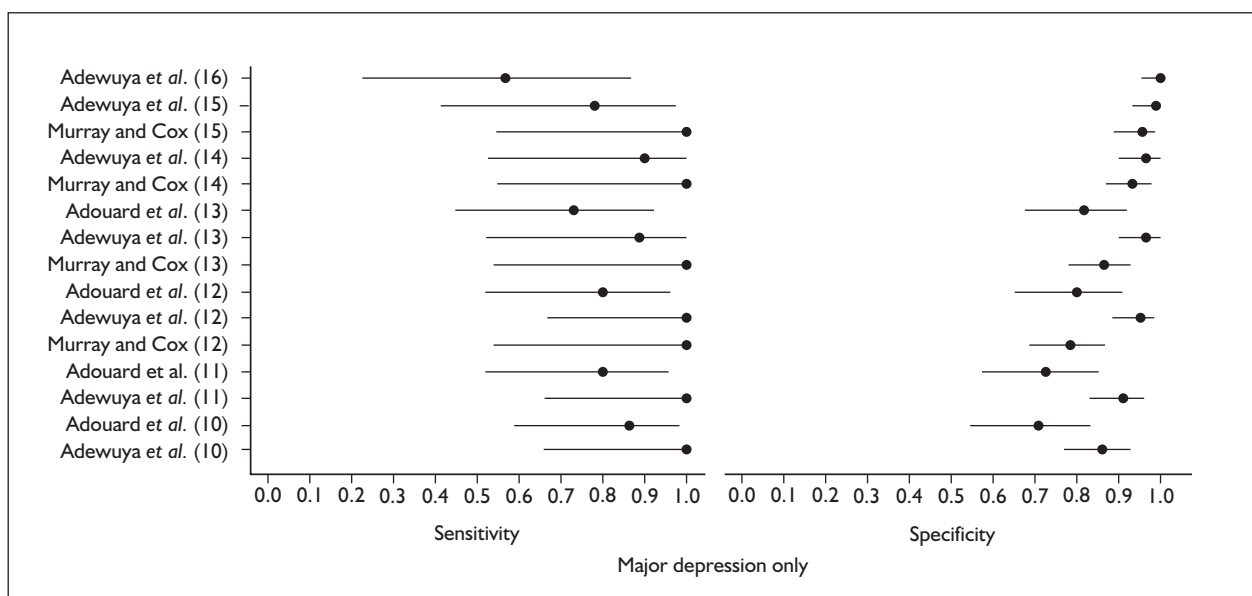
FIGURE 6 *Validation studies using the EPDS prenatally to diagnose major depression. Values in parentheses on the left-hand side of the graph represent the various cut points used to define major depression.*
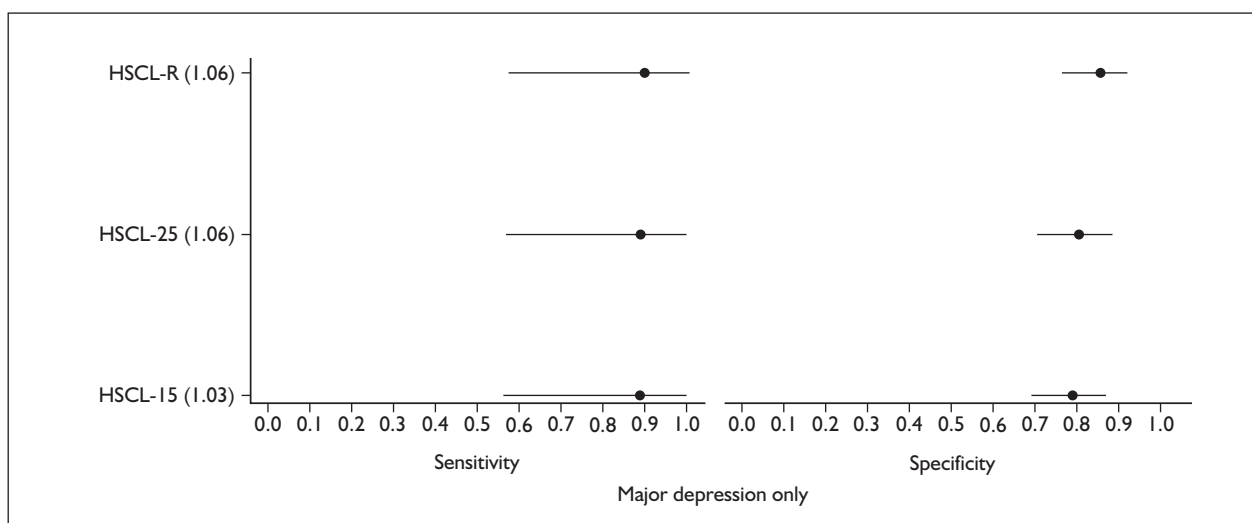

FIGURE 7 *Single validation study using the HSCL prenatally. Values in parentheses on the left-hand side of the graph represent the various cut points used to define major depression.*

to 1.00 (specificity 0.32). For the study using the Zung's SDS women were examined twice during the prenatal period – during the first trimester and during the third trimester. Sensitivity and specificity were reported for a range of cut points for both time points. Sensitivity ranged from 0.46 (specificity 0.93) to 1.00 (specificity 0.03) during the first trimester and from 0.10 (specificity 0.97) to 1.00 (specificity 0.33) during the third trimester. The authors set an optimal cut point of 23, giving a sensitivity of 0.91 (specificity 0.70) during the first trimester and a sensitivity of 0.70 (specificity

0.76) during the third trimester. The single study utilising the HADS reported a single cut point of 8 giving a sensitivity of 0.90 and a specificity of 0.91.

From *Figure 10* we can see that, from the studies identified, the EPDS had, on average, higher sensitivity and specificity values than those for the Zung's SDS, irrespective of the cut point used. It is also clear that the single study using HADS demonstrated relatively high sensitivity and specificity values compared with the other two instruments.
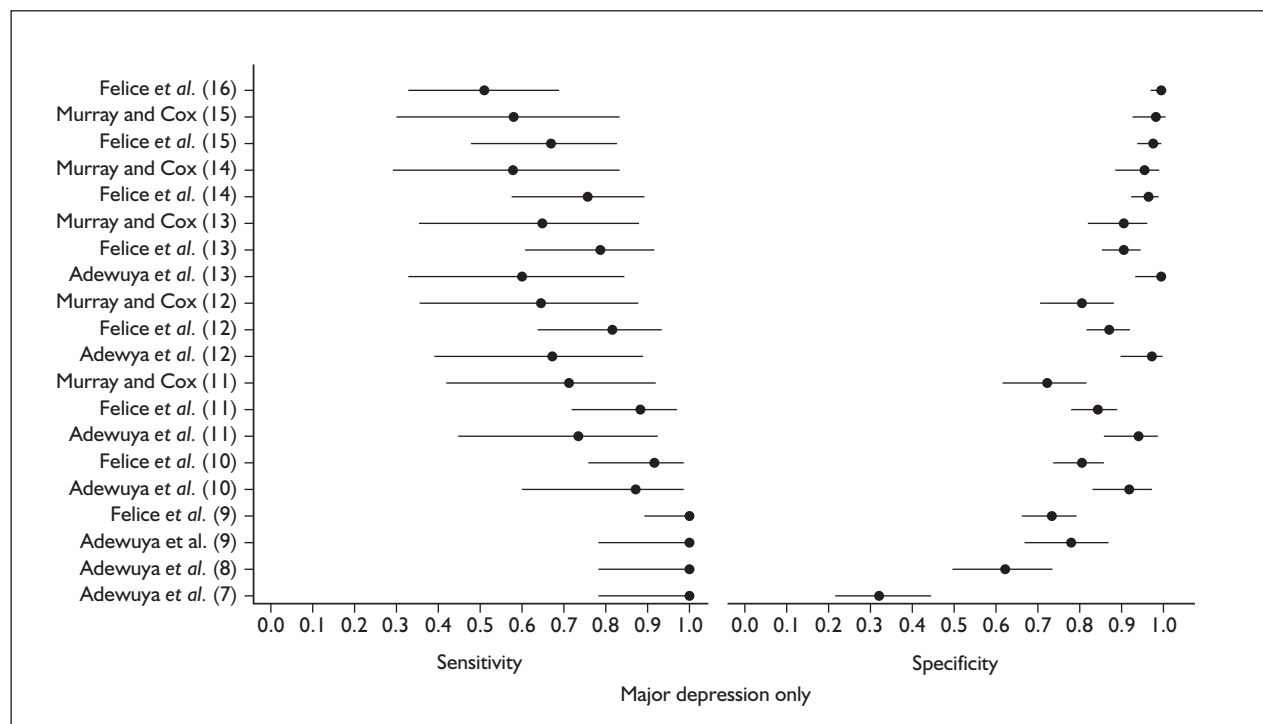
31

FIGURE 8 *Studies using the EPDS prenatally to diagnose major or minor depression. Values in parentheses on the left-hand side of the graph represent the various cut points used to define major or minor depression.*

Any psychiatric disorder

One study[113] administered the BDI during pregnancy (32 weeks' gestation) to identify women with current or remitting depression. Sensitivities and specificities were reported for a range of cut points (*Figure 11*). Sensitivities ranged from 0.67 (specificity 0.96) to 1.00 (specificity 0.68). The author recommended a cut point of 16 giving a sensitivity of 0.83 and a specificity of 0.89.

Other classifications

Two studies[108,115] used the General Health Questionnaire (GHQ) to identify other types of disorders. Pooling was not undertaken as each study classified different disorders. The first study[108] focused on identifying cases of anxiety and depression. Sensitivity and specificity were reported for a single cut point of 3 and were found to be 0.83 and 0.81 respectively. The second study[115] focused on identifying all RDC diagnoses. Women were examined twice during the prenatal period – in the first trimester and in the third trimester. Sensitivity and specificity were reported for a range of cut points for both time points. Sensitivity ranged from 0.28 (specificity 0.93) to 1 (specificity 0.03) during the first trimester and from 0.08 (specificity 0.95) to 0.92 (specificity 0.13) during the third trimester. The authors set an optimal cut point of 8, giving a sensitivity of 0.83 (specificity 0.71) during the first trimester and a sensitivity of 0.39 (specificity 0.82) during the third trimester.

Postnatal results

This section focuses on 55 studies (10,651 women) meeting the inclusion criteria that administered identification strategies in the postnatal period.

Quality assessment

Quality assessment was undertaken using QUADAS and the results are shown in *Figure 12*. There was variability in the results of the quality assessment. Studies demonstrated high quality in five out of the eight questions focusing on bias (questions 3, 5, 6, 7 and 14), with over 70% of studies scoring 'yes' in answer to these questions. Furthermore, for three questions (questions 3, 6 and 7) 100% of studies scored 'yes'. Scores were also favourable for the three questions relating to reporting quality (questions 8, 9 and 13); over 80% of studies scored 'yes' in answer to these questions.

Several QUADAS items were poorly described in the diagnostic studies. This included one of the questions relating to variability: in 56% of studies it was unclear how participants were selected for the study. The other question relating to variability (question 1) performed much better with 78% of studies scoring 'yes' in answer to this question. Question 10 had the lowest quality rating. It was unclear from 78% of the studies whether the index tests were interpreted without knowledge of the reference standard. Interestingly, the studies performed better on the related question 11;
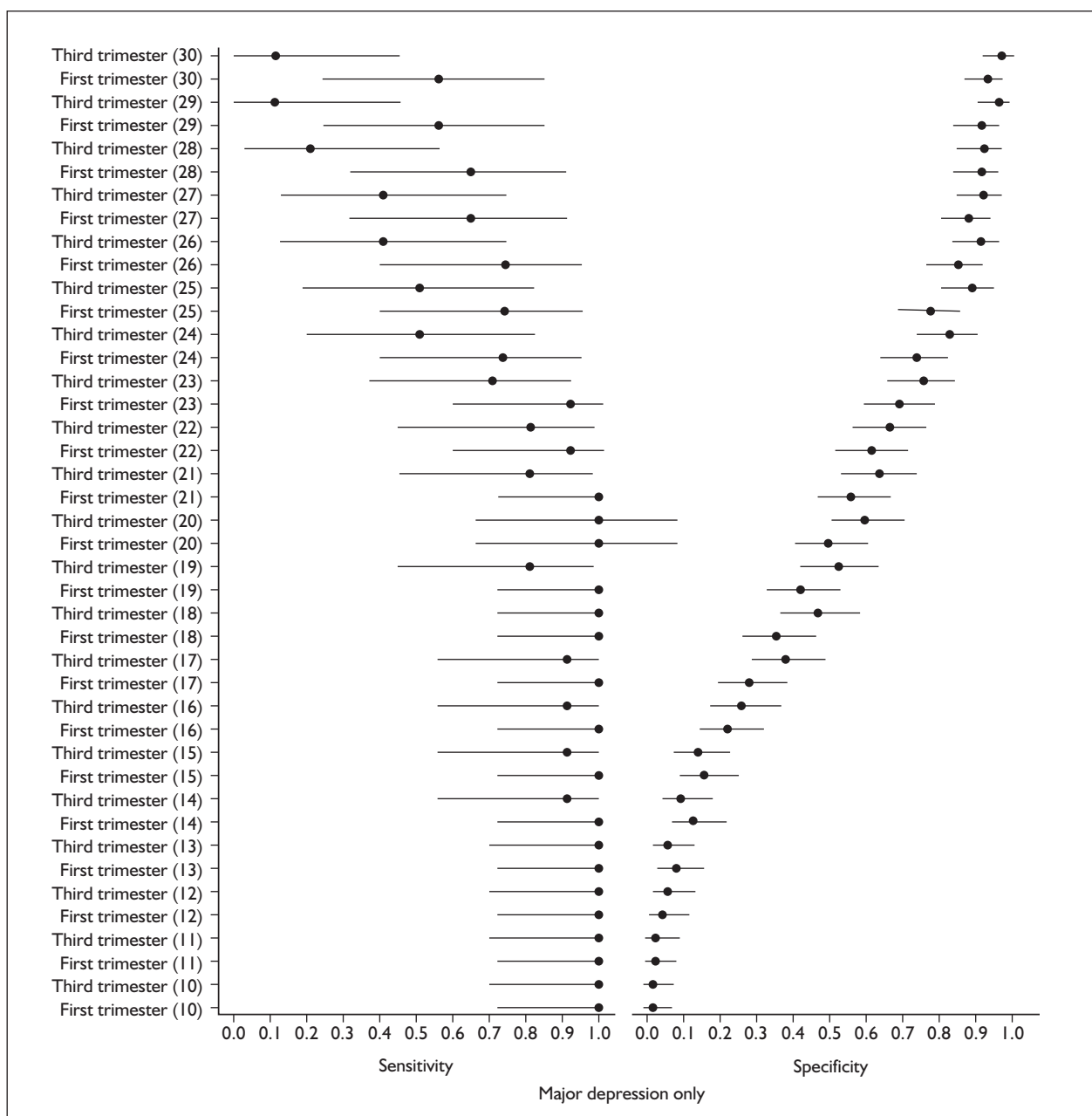
FIGURE 9 *Single study using Zung's Self-rating Depression Scale prenatally. Values on the left-hand side of the graph represent the time of administration (first or third trimester) and the values in parentheses display the various cut points used to define major or minor depression.*

for this question it was unclear in 38% of studies whether the reference standards were interpreted without knowledge of the index test. Finally, the question focusing on disease progression bias also scored poorly with only 53% of studies administering the index test and reference standard within 2 weeks of each other. It was unclear in the majority of studies when instruments were actually administered.

Characteristics of included studies

Studies were published between 1987 and 2007 and were undertaken in a variety of countries: eight in the UK,[16,33,34,36,117–120] four in Australia,[32,35,121,122] three each in Canada,[37,123,124] Japan,[106,115,125] Nigeria[126–128] and Spain,[103,129,130] two each in Austria,[105,131] France,[132,133] Italy,[134,135] Norway,[136,137] Sweden,[138,139] Thailand,[140,141] Turkey[142,143] and the USA,[31,44,144] and single studies in Chile,[145] Hong Kong,[146–148] Malaysia,[149,150] Malta,[116] Morocco,[151] Nepal,[152] New Zealand,[153] Peru,[104] South Africa,[154] Taiwan[155] and United Arab Emirates.[156] The percentage of women with PND in these studies ranged from 0.3% to 76% according to the reference standard criteria. In total, 13 different instruments were reported as being
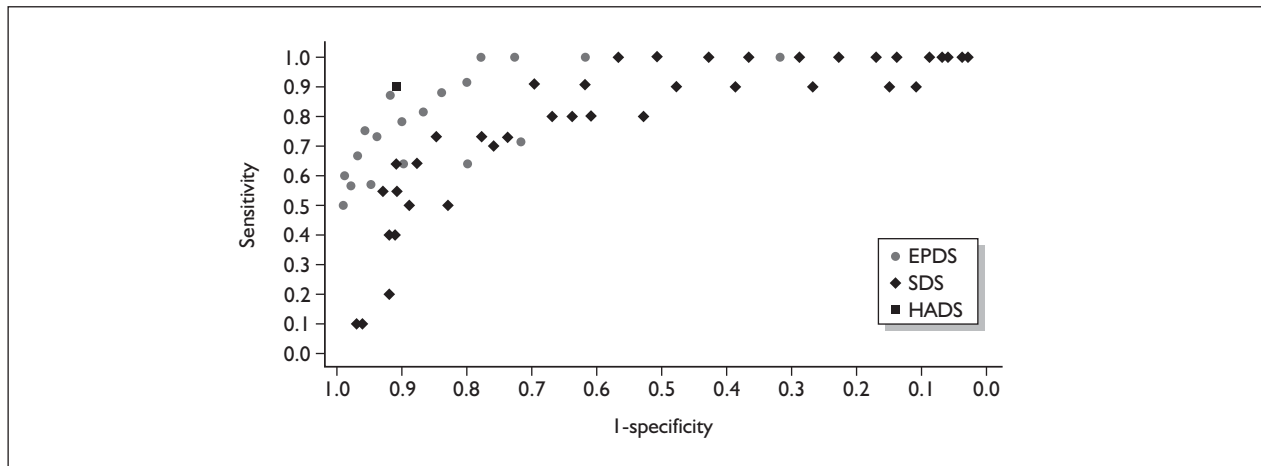
FIGURE 10 *Summary of identification strategies used prenatally at varying cut points to diagnose major or minor depression.*
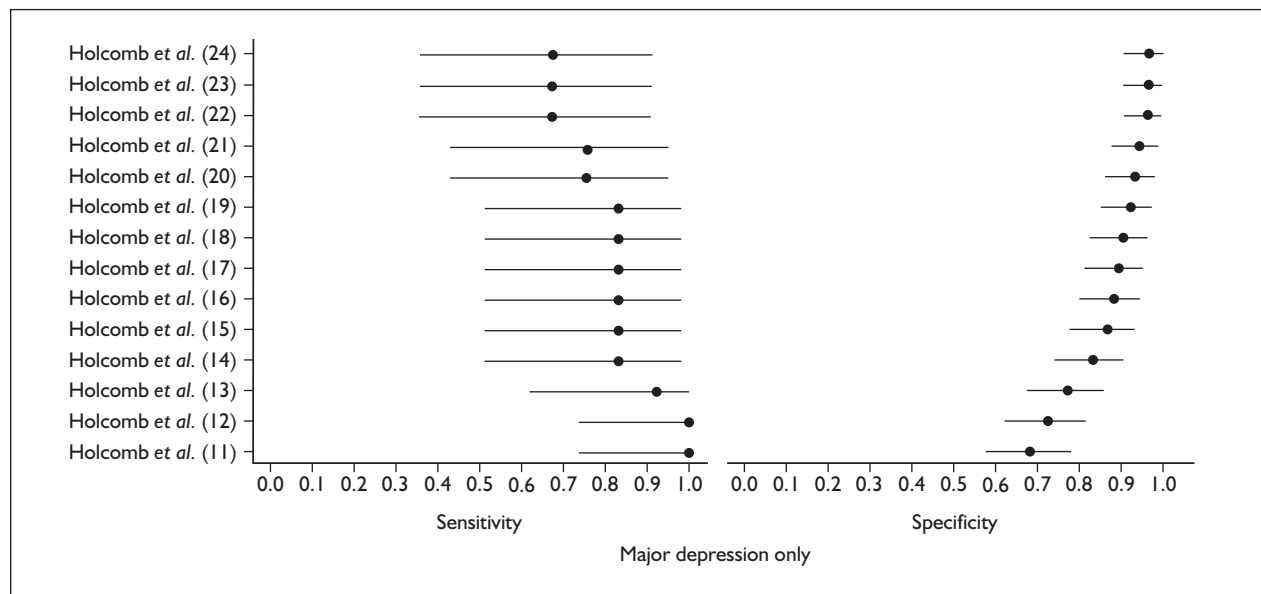

FIGURE 11 *Single study using the BDI prenatally. Values in parentheses on the left-hand side of the graph represent the various cut points used to define major or minor depression.*

used postnatally across the diagnostic accuracy studies included in this part of the review. *Figure 13* shows the number of studies utilising the different instruments. It is clear that the EPDS was the most frequently used instrument. The other identification strategies were used in two or fewer studies and included the Zung's Self-rating Depression Scale (SDS) ($n = 2$), HRSD ($n = 1$), MADRS ($n = 1$), Raskin ($n = 1$), SCL-90-R ($n = 1$) and EPDS–GHQ double test ($n = 1$). A number of the instruments have been translated into other languages and validated. For example, the EPDS has been translated from English into 18 other languages (*Table 9*). As previously stated there are multiple reference standards that can be used to establish whether an individual may be depressed or not and the type of depression that they have. In this part of the review, 29 (57%) of the studies used

DSM, 12 (24%) used RDC classifications, six (12%) used ICD, and three (6%) used Bedford College or Catego classifications. In the final study it was unclear which criteria were used; it was reported only that interviews were undertaken using SCID.

Major depression (DSM or equivalent) only

In total, 21 studies were included that compared identification strategies with a gold standard in the postnatal period for the diagnosis of major depression. Eight identification strategies were used: EPDS ($n = 18$), PDSS ($n = 3$), BDI ($n = 2$), BDI-II ($n = 2$), MADRS ($n = 1$), Raskin ($n = 1$), SCL-90-R depression ($n = 1$) and Zung's SDS ($n = 1$). Insufficient data were available to permit pooling for the majority of identification strategies. Only the EPDS had sufficient data to pool at a variety of
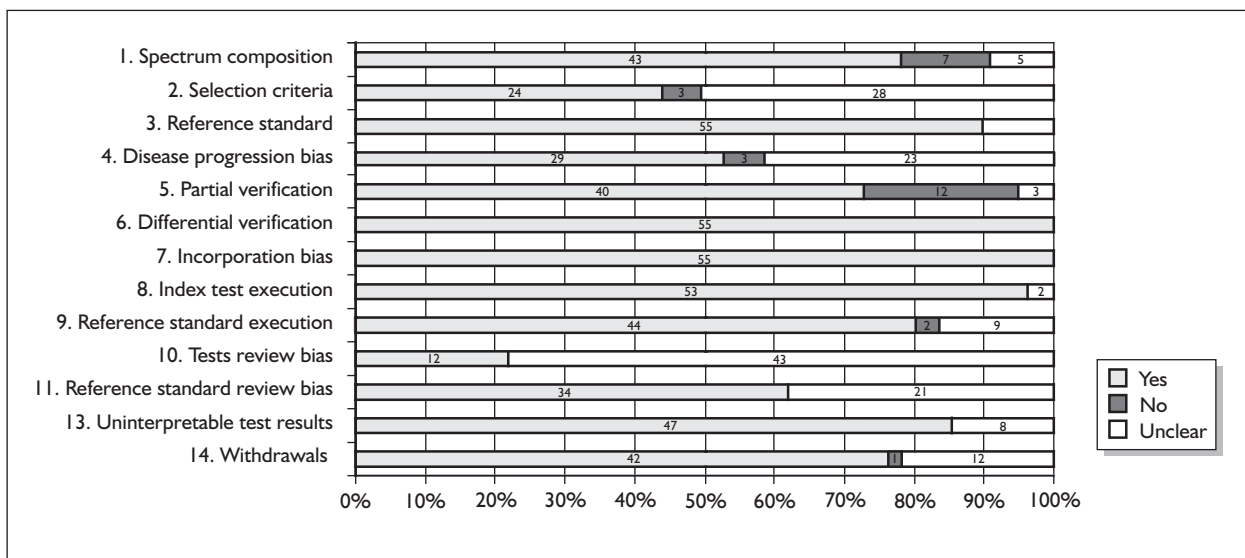
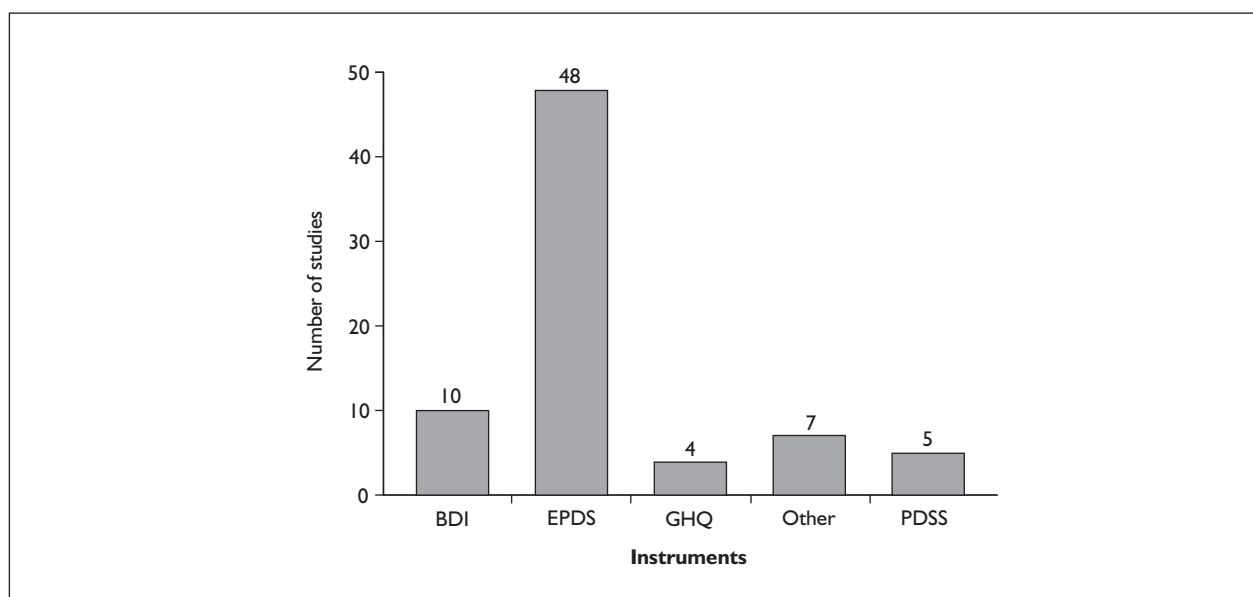**FIGURE 12** *Summary of quality assessment for postnatal studies.*

FIGURE 13 *Number of included studies using different instruments postnatally.*

cut points (7–16). A summary of the sensitivity and specificity of the studies using all of the instruments is shown in *Figure 14*.

When the studies were combined the pooled sensitivities ranged from 0.60 (95% CI 0.47 to 0.71) to 0.96 (95% CI 0.90 to 0.98) and specificities ranged from 0.45 (95% CI 0.26 to 0.66) to 0.97 (95% CI 0.92 to 0.99) for the diagnosis of major depression at a range of cut points using the EPDS (*Table 10*). The optimal cut point, in terms of the trade-offs between sensitivity and specificity (i.e. Youden's index), was 12 (*Figure 15*). At this cut

point a pooled sensitivity of 0.86 (95% CI 0.81 to 0.89) and specificity of 0.87 (95% CI 0.80 to 0.92) were derived. The positive likelihood ratio associated with the sensitivity and specificity was 6.66 (95% CI 4.32 to 10.28) and the pooled DOR was 40.54 (95% CI 24.22 to 67.88). A summary plot of sensitivity and specificity in ROC space, summarising each study at cut point 12, weighted by study size can be seen in *Figure 16*; additional plots for the other cut points can be found in Appendix 4. Although all of the studies used the EPDS, used a widely recognised gold standard and were focusing on identifying major depression the

TABLE 9 Characteristics of included studies using identification strategies postnatally

Study	Version	Study sample, validation sample	Identification strategy	Time of admin., setting if different	Interview type, gold standard	Type of depression	Cases (%)	Sample size	Mean age (years)
Abiodun, 2006,[126] Nigeria	English or Yoruba	Postnatal clinics at three primary health-care centres, all sample	EPDS	6 weeks postnatally	PSE, ICD-10	Any psychiatric disorder	67 (18.6)	360	27.9
Adewuya et al., 2005,[128] Morocco	English or Yoruba	Postnatal clinics and immunisation clinics at five health centres, EPDS ≥ 9 + BDI ≥ 10 + random sample with EPDS < 9 + BDI < 10	EPDS, BDI	6 weeks postnatally	SCID-NP, DSM-III-R	Major or minor	128 (14.6)	876	28.8
Agoub et al., 2005,[151] Nigeria	Arabic	Maternal and infant health unit in a primary health-care setting, all sample	EPDS	2–3 weeks postnatally	MINI, DSM-IV	Major only	27 (18.8)	144	30.3
Aslan et al., 1997,[143] Turkey	Turkish	Primary health-care centre, unclear	EPDS	Unclear	SCID-NP, unclear	Any psychiatric disorder	19 (25.0)	76	–
Aydin et al., 2004,[142] Turkey	Turkish	Women attending primary health-care clinics, all sample	EPDS	Within 1 year postnatally	SCID-I, DSM-IV	Major or minor	49 (14.4)	341	26.6
Bagedahl-Strindlund et al., 1998,[138] Sweden	Swedish	All Swedish speaking mothers visiting 15 Well Baby Clinics, EPDS > 12	EPDS	3 months postnatally, mean = 13.5 weeks	Unclear, RDC	Major only Major or minor	1 (0.3) 19 (6.3)	303 –	28.2 –
Barnett et al., 1999,[122] Australia	Vietnamese Anglo-Celtic Arabic	Four antenatal clinics, all sample	EPDS	6 weeks postnatally, at home	DIS, DSM-III-R	Major only	5 (4.4) 7 (6.7) 9 (9.5)	113 105 95	27.0 – –
Beattie-Clarke, 2003,[124] Canada	English	Advertisements and women attending prenatal, postnatal and parenting groups, all sample	EPDS, PDSS, BDI-II	1–12 months postnatally: 1–2 months, n = 18; 2–4, n = 31; 4–6, n = 12; 6–9, n = 14; 9–12, n = 27	SCID, DSM-IV	Major only Major or minor	17 (16.5) 25 (24.3)	103 –	23.8 –
Beck and Gable, 2001,[31] USA; Beck and Gable, 2001[44] USA	English	Preparation for childbirth classes or newspaper advert, all sample	PDSS, EPDS, BDI-II	2–12 weeks postnatally, mean = 39 days	SCID, DSM-IV	Major only Major or minor	18 (14.8) 46 (30.7)	122 150	31.0 –

Study	Version	Study sample, validation sample	Identification strategy	Time of admin., setting if different	Interview type, gold standard	Type of depression	Cases (%)	Sample size	Mean age (years)
Beck and Gable, 2005,[144] USA	Spanish	Acute care, community teaching hospital and perinatal clinic, all sample	PDSS-7, PDSS-35	2–12 weeks postnatally, mean = 27 days	SCID, DSM-IV	Major or minor	55 (36.7)	150	25.8
Benvenuti et al., 1999,[134] Italy	Italian	Obstetric clinic at large university hospital, all sample	EPDS	8–12 weeks postnatally	MINI, DSM-III-R	Any psychiatric disorder	18 (15.9)	113	31.9
Bergant et al., 1998,[105] Austria	German	Randomly selected women after childbirth in university hospital, all sample	EPDS	4 days postnatally	Unclear; ICD-10	Major or minor	22 (20.0)	110	28.6
Berle et al., 2003,[136] Norway	Norwegian	Women attending routine postnatal visits, EPDS ≥ 8 + every tenth EPDS < 8	EPDS	6–12 weeks postnatally	MINI, DSM-IV	Major only Major or minor	27 (27.0) 41 (41.0)	100 –	30.3 –
Boyce et al., 1993,[32] Australia	English	Mothers' Advisory Clinics (baby health clinics staffed by community nurses) + women referred to hospital psychiatric department for outpatient treatment of PND, all sample	EPDS	Within 6 months postnatally, mean = 12 weeks postnatally at home	DIS, DSM-III-R at home	Major only	9 (8.7)	103	28.4
Carpiniello et al., 1997,[135] Italy	Italian	Consecutive admissions for delivery at obstetric clinic in university hospital, all sample	EPDS	4–6 weeks postnatally, at home	PSE, ICD-Catego at home	Any psychiatric disorder	9 (14.8)	61	31.6
Cox et al., 1987,[16] UK	English	Part of larger study, health visitors identified women as potentially depressed + 12 non-depressed women, all sample	EPDS	3 months postnatally, at home	SPI, RDC at home or clinic	Major or minor	35 (41.7)	84	26.0
Cox et al., 1996,[120] UK	English	Part of controlled study including non-postnatal women, recruited from GP age/sex registers mailed EPDS, EPDS ≥ 9 + one-third EPDS < 9	EPDS	6 months postnatally, at home	CIS, RDC at home	Major only Major or minor	8 (6.2) 21 (16.3)	128 –	26.3 –

TABLE 9 *Characteristics of included studies using identification strategies postnatally (continued)*

Study	Version	Study sample, validation sample	Identification strategy	Time of admin., setting if different	Interview type, gold standard	Type of depression	Cases (%)	Sample size	Mean age (years)
Eberhard-Gran et al., 2001,[137] Norway	Norwegian	Women from a community-based questionnaire study, EPDS ≥ 10 + a control group; control group women included those with EPDS < 10 and who delivered closest in time to the high scoring woman	EPDS	10 weeks postnatally, local primary health-care centre	PRIME-MD, DSM-IV	Major only	9 (16.1)	56	–
						Major or minor	21 (37.5)	–	–
Felice et al., 2006,[116a] Malta	Maltese	Antenatal clinic at hospital, all sample	EPDS	8 weeks postnatally, home	CIS-R, ICD-10	Major or minor	18 (8.1)	223	27.1
Garcia-Esteve et al., 2003,[129] Spain	Spanish	Maternity hospital, EPDS ≥ 9 + 10% random EPDS < 9	EPDS	6 weeks postnatally	SCID, DSM-III-R	Major only	36 (3.2)	1123	30.0
						Major or minor	100 (8.9)	–	–
Ghubash et al., 1997,[156] United Arab Emirates	Arabic	Postnatal ward at hospital, all sample	EPDS	1 week postnatally	PSE, Catego, 8 weeks postnatally	Any psychiatric disorder	13 (13.7)	95	28.6
Guedeney and Fermanian, 1998,[133] France	French	Random sample of women involved with public service mainly staffed with community nurses, provides home visits, consultations in baby health clinics + identified as probably depressed by nurses, all sample	EPDS	Within first 4 months postnatally, mean = 7 weeks at home	PSE, RDC	Major or minor	45 (51.7)	87	30.4
Harris et al., 1989,[33] UK	English	Part of larger population study for postnatal thyroid dysfunction, all sample	BDI	6 weeks postnatally, routine postnatal follow-up clinic	Unclear; DSM-III	Major only	19 (14.7)	129	24.6
			Raskin				22 (15.0)	146	–
			MADRS				22 (15.1)	147	–
			EPDS				22 (17.5)	126	–
Holt, 1995,[153] New Zealand	English	All women attending 6-week postnatal examinations, all sample	EPDS	6 weeks postnatally	Review of case notes, DSM-III	Major only	7 (5.8)	121	26.0
Jadresic et al., 1995,[145] Chile	Spanish	Antenatal clinic at university hospital, all sample	EPDS	2–3 months postnatally	PAS, RDC	Major or minor	11 (10.2)	108	27.7

Study	Version	Study sample, validation sample	Identification strategy	Time of admin., setting if different	Interview type, gold standard	Type of depression	Cases (%)	Sample size	Mean age (years)
Jardi et al., 2006,[132] France	French	Maternity unit at university hospital, EPDS > 8 + same number randomly selected EPDS ≤ 8	EPDS	3–5 days postnatally	MINI, DSM-IV, telephone at 8 weeks postnatally	Major or minor	112 (30.8)	364	28.8
Kitamura et al., 1994,[115a] Japan	Japanese	Volunteers participating in follow-up study, attending obstetric department at general hospital, all sample	Zung's SDS	5 days postnatally; 1 month postnatally	SADS, RDC	Major or minor	11 (12.1); 9 (8.9)	91; 101	28.0; –
			GHQ-30	5 days postnatally; 1 month postnatally		Other	18 (20.0); 16 (16.2)	90; 99	–; –
Lawrie et al., 1998,[154] South Africa	Multiple: English, African, Zulu, Tswana, Sotho, Xhosa + others	Mother and child academic hospital serving only women who experienced obstetric complications, required Caesarean section or requested sterilisation for family planning – women seen at 6 weeks, all sample	EPDS	6 weeks postnatally, EPDS read to women	Unclear, DSM-IV	Major only; Major or minor	8 (7.8); 25 (24.5)	102; –	28.1; –
Lee et al., 2000,[147] Hong Kong; Lee et al., 2001,[146] Hong Kong; Lee et al., 2003,[148] Hong Kong	Chinese	Consecutive admissions to postnatal ward at university-affiliated general hospital, all sample	EPDS, GHQ-12, BDI, EPDS–GHQ double test; EPDS, GHQ-12, BDI	6 weeks postnatally; 2 days postnatally	SCID-NP, DSM-III-R at 6 weeks postnatally	Major or minor	17 (11.7)	145	29.0
Leverton and Elliott, 2000,[34] UK	English	Antenatal clinic at hospital for larger prevention study, all sample	EPDS	3 months postnatally, at home	PSE, Catego (last 2 weeks of month); PSE, Catego (any time in last month); PSE, Bedford (last 2 weeks of month); PSE, Bedford (any time in last month)	Other	9 (4.5); 10 (5.0); 14 (7); 16 (8.0)	199; –; –; –	–; –; –; –

continued

TABLE 9 *Characteristics of included studies using identification strategies postnatally (continued)*

Study	Version	Study sample, validation sample	Identification strategy	Time of admin., setting if different	Interview type, gold standard	Type of depression	Cases (%)	Sample size	Mean age (years)
Mahmud et al., 2003,[150] Malaysia; Mahmud et al., 2004,[149] Malaysia	Malay	Women attending health centre, all sample	EPDS	4–12 weeks postnatally, mean = 55.5 days	CIS, ICD-10	Major or minor	9 (14.1)	64	28.7
			BDI-II	4–12 weeks postnatally, mean = 55.1 days			7 (11.5)	61	26.6
Matthey et al., 2001,[35] Australia	English	Couples attending Preparation for Parenthood classes in hospital in Australia, all sample	EPDS	6–7 weeks postnatally, at home	DIS, DSM-IV 3 days after completion	Other / Major or minor	37 (15.5) / 24 (10.4)	238 / 230	27.2 / –
Milgrom et al., 2005,[121] Australia	English	Women attending 47 maternal and child health centres in Australia, EPDS ≥ 12	EPDS, BDI	4 months postnatally	CIDI, DSM-IV, timing between screening and diagnosis averaged 1–2 weeks	Other	260 (75.6)	344	30.1
Murray and Carothers, 1990,[36] UK	English	Postnatal wards at maternity hospitals approached for other study, EPDS ≥ 13 + random sample EPDS 10–12 + 1 in 10 of EPDS < 10	EPDS	6 weeks postnatally, postal	SPI, RDC	Major only / Major or minor	Values estimated by logistic regression	702	–
Muzik et al., 2000,[131] Austria	German	Drawn from larger epidemiological study of PND in Austria, EPDS > 7	EPDS, SDS, SCL-90-R depression	3 or 6 months postnatally	SCID, DSM-III-R	Major only	9 (18.0)	50	28.0
Navarro et al., 2007,[130] Spain	Spanish	Routine postnatal follow-up at acute care teaching hospital, random sample 15% EPDS <7, 50% EPDS between 7 and 9, 60% EPDS between 10 and 12, 100% EPDS ≥ 13	GHQ-12, EPDS	6 weeks postnatally	SCID-NP, DSM-IV	Other	180 (44.4)	405	–
Okano et al., 1996,[106] Japan	Japanese	Antenatal clinic in the later part of pregnancy, all sample	EPDS	1 month postnatally	SADS, RDC	Major or minor	4 (8.5)	47	–

Study	Version	Study sample, validation sample	Identification strategy	Time of admin., setting if different	Interview type, gold standard	Type of depression	Cases (%)	Sample size	Mean age (years)
Pitanupong et al., 2007,[141] Thailand	Thai	Women giving birth and receiving follow-up care at university hospital, all sample	EPDS	6–8 weeks postnatally	Unclear, DSM-IV	Major or minor	38 (10.8)	351	27.9
Regmi et al., 2002,[152] Nepal	Unclear	Immunisation visit at postnatal clinic at university hospital, EPDS ≥ 13 + every fifth EPDS < 13	EPDS	2–3 months postnatally	SCID, DSM-IV	Major only	5 (5.0)	100	–
Teng et al., 2005,[155] Taiwan	Taiwanese	Maternity wards at university hospital, all sample	EPDS, BDI-II	6 weeks postnatally	MINI, DSM-IV	Any psychiatric disorder	24 (11.8)	203	29.0
Terren et al., 2003,[103] Spain	Spanish	Women attending postnatal clinic at university hospital, EPDS ≥ 9 + random 16% sample with EPDS < 9	EPDS	6 weeks postnatally	SCID-NP, DSM-IV	Major or minor	100 (29.9)	334	30.1
Thompson et al., 1998,[118] UK	English	Consecutive women presenting at 16 weeks' gestation at hospital, 50% GHQ-30 ≥ 7 + ~50% positive thyroid antibody	EPDS	8 weeks postnatally	Unclear, RDC	Major or minor	8 (7.3)	109	–
				12 weeks postnatally			17 (11.4)	149	–
				20 weeks postnatally			18 (12.8)	141	–
				28 weeks postnatally			15 (10.4)	144	–
			HAMD	8 weeks postnatally			10 (9.3)	108	–
				12 weeks postnatally			15 (10.6)	141	–
				20 weeks postnatally			12 (8.9)	135	–
				28 weeks postnatally			13 (9.6)	135	–
Uwakwe and Okonkwo, 2003,[127] Nigeria	English or Igbo	Wards and postnatal clinics at university hospital, all sample	EPDS	6–8 weeks postnatally, 98% within 6 weeks	CIDI adapted, ICD-10, SCL within 48 hours of screening	Any psychiatric disorder	24 (10.7)	225	28.9
Vega-Dienstmaier et al., 2002,[104] Peru	Spanish	Family planning centres and routine paediatric follow-up consultations, all sample	EPDS	Within 1 year	SCID, DSM-IV	Major only	19 (5.9)	321	25.1

TABLE 9 *Characteristics of included studies using identification strategies postnatally (continued)*

Study	Version	Study sample, validation sample	Identification strategy	Time of admin., setting if different	Interview type, gold standard	Type of depression	Cases (%)	Sample size	Mean age (years)
Vittayanont et al., 2006,[140] Thailand	Thai	University hospital, all sample	PDSS	6–8 weeks postnatally	Unclear, DSM-IV	Major only Major or minor	4 (1.0) 40 (10.0)	400	27.9
Werrett and Clifford, 2006,[117] UK	Punjabi and English	Convenience sample of bilingual mothers from caseloads of health visitors, all sample	EPDS	5–8 weeks postnatally	CIDI, ICD-10, 1 week later at home or health centre	Major or minor	7 (30.4)	23	28.6
Whiffen, 1988,[123] Canada	English	Public health prenatal classes, all sample	BDI	6–8 weeks postnatally, mean = 45 days	SADS, RDC interview, average 7.6 days after BDI	Major only Major or minor	9 (7.5) 21 (17.5)	120 –	28.0 –
Wickberg and Hwang, 1996,[139] Sweden	Swedish	17 child health clinics, EPDS 12 both times + random sample (n = 16) EPDS 10 or 11 + (n = 21) EPDS < 10	EPDS	Twice at 2 and 3 months postnatally	MADRS, DSM-III-R, 1–2 weeks after EPDS at home	Major only	56 (43.8)	128	28.1
Yamashita et al., 2000,[125] Japan	Japanese	Mothers admitted for delivery to maternity ward of university hospital, all sample	EPDS	1 month postnatally, routine 1-month postnatal check	SADS, RDC, telephone at 3 weeks postnatally	Major or minor	11 (14.7)	75	30.5
Yoshida et al., 1997,[119] UK	Japanese	Antenatal classes for pregnant Japanese women and advertisement in maternity guidebook for pregnant Japanese women in the UK, all sample	EPDS	1 month postnatally, postal	SADS, RDC at home, 3 months postnatally	Major or minor	8 (8.2)	97	30.0
Zelkowitz and Milet, 1995,[37] Canada	English	Two community health-care centres, unclear	EPDS	6–8 weeks postnatally, telephone	SCID-NP, DSM-III-R, within 10 days of telephone screening	Any psychiatric disorder	43 (48.3)	89	30.5

BDI, Beck Depression Inventory; CIDI, Composite International Diagnostic Interview; CIS, Clinical Interview Schedule; DIS, Diagnostic Interview Schedule; DSM, Diagnostic and Statistical Manual; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; HAMD, Hamilton Rating Scale for Depression; ICD, International Classification of Diseases; MADRS, Montgomery–Asberg Depression Rating Scale; MINI, Mini-International Neuropsychiatric Interview; PAS, Psychiatric Assessment Schedule; PDSS, Postpartum Depression Screening Scale; PND, post natal depression; PRIME-MD, Primary Care Evaluation of Mental Disorders; PSE, Present State Examination; RDC, Research Diagnostic Criteria; SADS, Schedule for Affective Disorders and Schizophrenia; SCID, Structured Clinical Interview for DSM; SCL-90-R, Symptom Checklist-90-Revised; Zung's SDS, Zung's Self-rating Depression Scale; SPI, Standardised Psychiatric Interview.
a Studies also validated identification strategies prenatally.
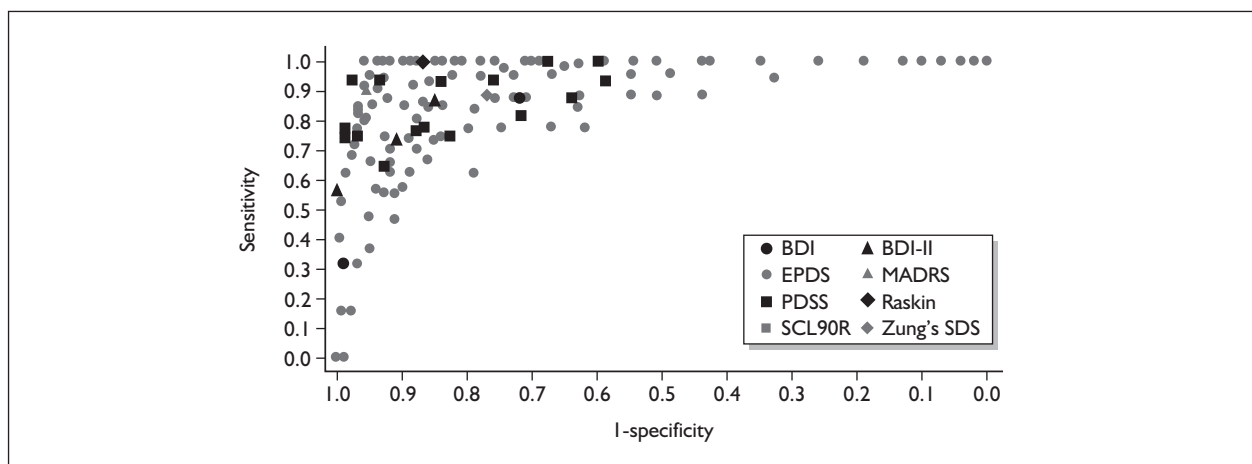
FIGURE 14 *Summary of identification strategies used postnatally at varying cut points to diagnose major depression.*

TABLE 10 *Summary of the sensitivity and specificity of the EPDS postnatally at varying cut points for major depression*

Cut point	Sensitivity	Specificity	LR+	LR–	DOR	*n*	I^2
7	0.96 (0.90 to 0.98)	0.45 (0.26 to 0.66)	1.76 (1.20 to 2.57)	0.09 (0.03 to 0.26)	19.89 (5.39 to 73.43)	4	0
8	0.94 (0.89 to 0.97)	0.58 (0.46 to 0.69)	2.23 (1.68 to 2.96)	0.10 (0.05 to 0.20)	21.45 (9.05 to 50.83)	9	0
9	0.92 (0.87 to 0.96)	0.66 (0.55 to 0.75)	2.71 (1.99 to 3.69)	0.12 (0.06 to 0.21)	23.52 (10.36 to 53.43)	10	21
10	0.92 (0.87 to 0.95)	0.77 (0.70 to 0.83)	4.00 (2.94 to 5.45)	0.11 (0.06 to 0.18)	37.17 (17.61 to 78.44)	14	38
11	0.87 (0.80 to 0.92)	0.84 (0.76 to 0.89)	5.35 (3.50 to 8.17)	0.15 (0.10 to 0.24)	34.91 (16.49 to 73.92)	12	59
12	0.86 (0.81 to 0.89)	0.87 (0.80 to 0.92)	6.66 (4.32 to 10.28)	0.16 (0.12 to 0.22)	40.54 (24.22 to 67.88)	15	63
13	0.79 (0.74 to 0.83)	0.89 (0.85 to 0.92)	7.50 (5.38 to 10.45)	0.23 (0.19 to 0.29)	32.29 (20.76 to 50.22)	18	86
14	0.73 (0.64 to 0.80)	0.92 (0.86 to 0.95)	9.04 (5.04 to 16.21)	0.30 (0.22 to 0.40)	30.37 (13.96 to 66.07)	8	89
15	0.65 (0.55 to 0.74)	0.96 (0.92 to 0.98)	15.10 (8.02 to 28.27)	0.36 (0.28 to 0.48)	41.41 (20.03 to 85.61)	7	88
16	0.60 (0.47 to 0.71)	0.97 (0.92 to 0.99)	22.46 (7.25 to 69.56)	0.42 (0.31 to 0.56)	54.04 (16.55 to 176.48)	4	93

DOR, diagnostic odds ratio; LR+, positive likelihood ratio; LR–, negative likelihood ratio.
Note: Values in parentheses are 95% confidence intervals.

I^2 value of 63% was high, indicating high between-study heterogeneity. Thus, care should be taken when interpreting the results.

When the psychometric attributes were pooled across the studies, a variety of levels of between-study heterogeneity were identified (range 0–93%). From *Table 10* we can see that moderate or high levels of between-study heterogeneity were identified when a cut point of 11 or higher was used. Meta-regression was undertaken for cut points 11, 12 and 13; because of the low number of studies reporting data for cut points above 13 we were unable to investigate any potential sources of heterogeneity.

Table 11 displays the results of the univariate analyses. As meta-regression was undertaken within each cut point, multiple tests were performed and therefore we must be cautious in interpreting the results. It can be seen that timing seemed to be an important factor across cut points 12 and 13. The
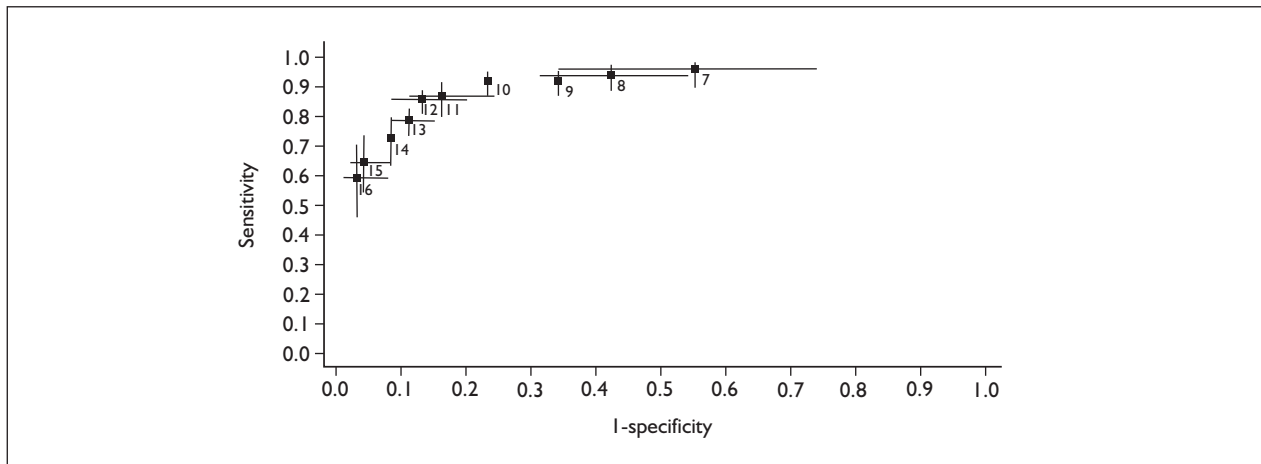
FIGURE 15 *Graphical summary of the sensitivity and specificity of the EPDS postnatally at varying cut points to diagnose major depression.*
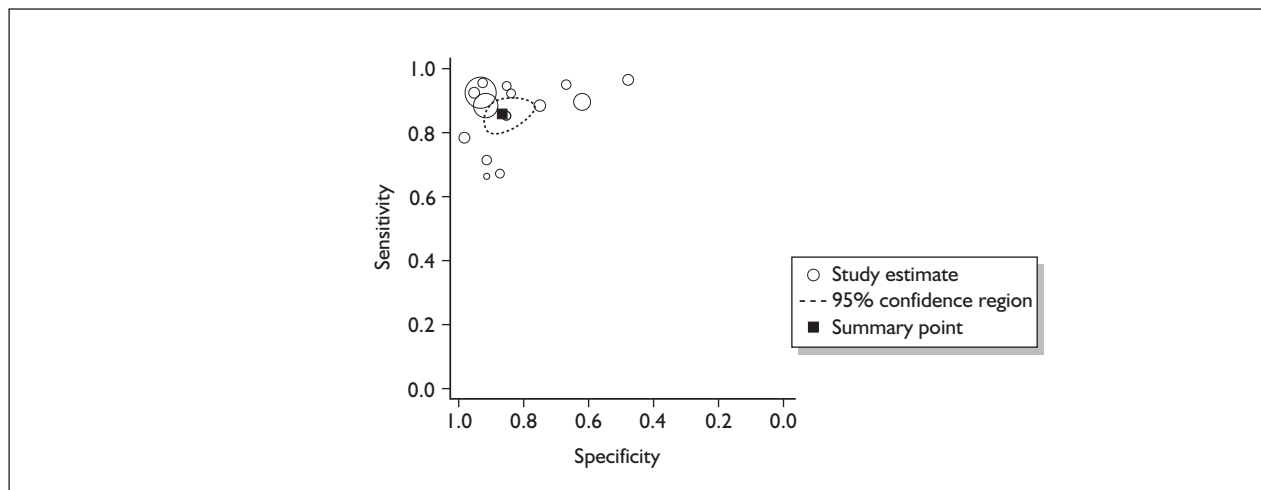
FIGURE 16 *EPDS SROC plot for diagnosis of major depression at cut point 12.*

DORs for studies administering the EPDS within 6 weeks postnatally were higher than those in studies administering the EPDS after 6 weeks postnatally (cut point 12: 84 versus 27; cut point 13: 68 versus 16). This indicates better discriminatory performance if the EPDS is administered within 6 weeks postnatally. The I^2 value reduced from 63% to 40% and 86% to 47% for cut points 12 and 13, respectively, when timing was included. For cut point 11, verification bias seemed to be the most important factor in explaining the heterogeneity in the DOR of the primary studies. The DOR for studies that performed a diagnostic interview on all women included in the study (or a random sample) was 26, whereas the DOR for studies that only performed a diagnostic interview on a non-random sample of women was 65. This analysis highlights that the DOR was inflated when a non-random sample of women was given a diagnostic interview, thus indicating the presence of verification bias.

Table 12 shows the results of the multivariate analyses. As meta-regression was undertaken within each cut point, multiple tests were performed, and, also, because of the low number of studies included we should be cautious when interpreting the results. When all variables were considered simultaneously, timing seemed to be the most important factor in explaining the heterogeneity in the DOR of the primary studies across all cut points, although none of the variables showed statistically significant differences at the $p \leq 0.05$ level. Across all cut points the I^2 value reduced, from 59% to 35%, 63% to 44% and 86% to 44% for cut points 11, 12 and 13 respectively. It is worth noting that the I^2 values in the multivariate analyses are not substantially reduced compared with the univariate analyses. As the I^2 values were above 0% there are obviously still some sources of variation that remain unexplained; however, the I^2 values did reduce and were classified as moderate.

TABLE 11 *Univariate meta-regression analyses*

Cut point	Predictor	DOR$_1$	DOR$_2$	Beta-coefficient (95% CI)	*p*-value	*I*^{2}
11	Prevalence	–	–	5.59 (–8.55 to 19.72)	0.40	62
	Verification bias	26.01	64.58	0.39 (0.08 to 1.98)	0.23	41
	Blinding	73.93	24.18	3.15 (0.66 to 15.10)	0.13	37
	Timing	56.81	21.29	2.08 (0.36 to 12.07)	0.38	53
12	Prevalence	–	–	–1.78 (–6.95 to 3.40)	0.47	59
	Verification bias	45.97	58.92	0.76 (0.22 to 2.70)	0.65	61
	Blinding	83.61	38.02	2.23 (0.68 to 7.36)	0.17	52
	Timing	83.78	26.68	3.14 (1.13 to 8.74)	0.03	40
13	Prevalence	–	–	–4.07 (–7.87 to –0.27)	0.04	58
	Verification bias	40.14	32.37	1.33 (0.30 to 2.92)	0.69	87
	Blinding	54.80	31.07	1.75 (0.45 to 6.86)	0.40	78
	Timing	68.48	15.67	4.10 (1.78 to 9.44)	0.002	47

CI, confidence interval; DOR, diagnostic odds ratio.

TABLE 12 *Multivariate meta-regression analyses*

Cut point	Predictor	Beta-coefficient (95% CI)	*p*-value	*I*^{2}
11	Prevalence	8.09 (–6.46 to 22.65)	0.23	35
	Verification bias	0.43 (0.01 to 20.08)	0.62	
	Blinding	1.13 (0.02 to 61.28)	0.94	
	Timing	3.48 (0.52 to 23.42)	0.17	
12	Prevalence	1.69 (–5.00 to 8.37)	0.59	44
	Verification bias	1.40 (0.15 to 13.10)	0.75	
	Blinding	2.42 (0.24 to 24.03)	0.41	
	Timing	3.09 (0.75 to 12.83)	0.11	
13	Prevalence	–1.06 (–5.81 to 3.70)	0.64	44
	Verification bias	1.61 (0.33 to 7.75)	0.52	
	Blinding	2.10 (0.41 to 10.71)	0.34	
	Timing	2.65 (0.87 to 8.04)	0.08	

Major or minor depression (DSM or equivalent)

In total, 35 studies compared identification strategies with a gold standard in the postnatal period for the diagnosis of major or minor depression. Eight identification strategies were used: EPDS (*n* = 28), PDSS (*n* = 4), BDI (*n* = 4), BDI-II (*n* = 3), GHQ-12 (*n* = 2), EPDS–GHQ (*n* = 1), Hamilton Rating Scale for Depression (HAMD) (*n* = 1), PDSS-short (*n* = 1) and Zung's SDS (*n* = 1). Insufficient data were available to permit pooling for the majority of identification strategies. The EPDS had sufficient data to pool at a variety of cut points (7–16), the BDI at cut point 10 and the HAMD at cut point 11. A summary of

the sensitivity and specificity of all of the studies is given in *Figure 17*. This shows that, from the studies identified, the EPDS, BDI, BDI-II and HAMD seem to have, on average, higher sensitivity and specificity values than the other identification strategies irrespective of the cut point used.

When the studies were combined the pooled sensitivities ranged from 0.31 (95% CI 0.19 to 0.47) to 0.91 (95% CI 0.80 to 0.96) and specificities ranged from 0.67 (95% CI 0.57 to 0.76) to 0.99 (95% CI 0.98 to 0.99) for the diagnosis of major or minor depression at a range of cut points using the EPDS (*Table 13*). The optimal cut point, in terms of the trade-offs between sensitivity and specificity
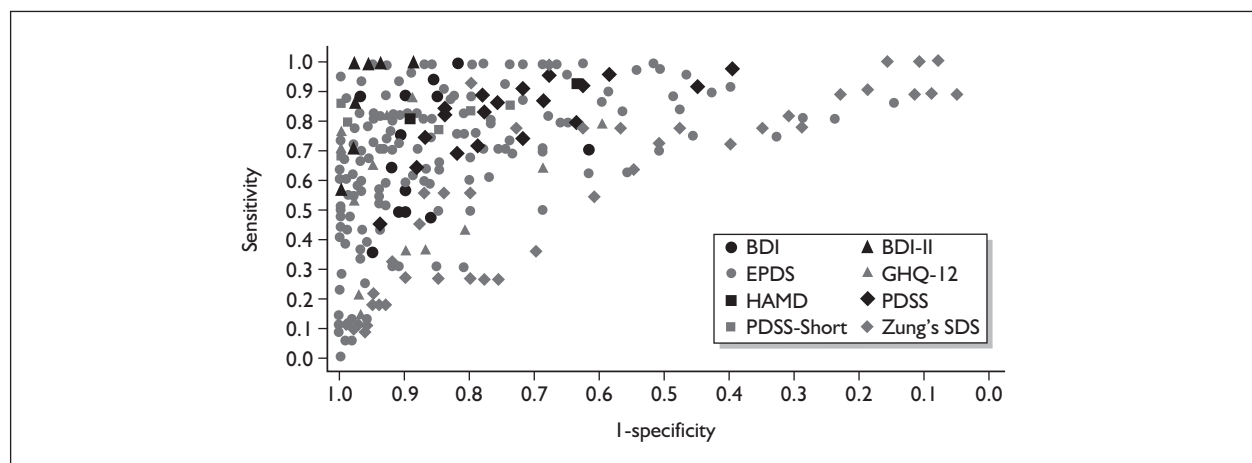
FIGURE 17 *Summary of identification strategies used postnatally at varying cut points to diagnose major or minor depression.*

TABLE 13 *Summary of the sensitivity and specificity of the EPDS postnatally at varying cut points for major or minor depression*

Cut point	Sensitivity	Specificity	LR+	LR–	DOR	n	I²
7	0.91 (0.80 to 0.96)	0.67 (0.57 to 0.76)	2.76 (2.07 to 3.68)	0.13 (0.05 to 0.32)	20.95 (7.65 to 27.34)	10	72
8	0.91 (0.84 to 0.95)	0.75 (0.64 to 0.83)	3.58 (2.46 to 5.23)	0.12 (0.06 to 0.22)	30.34 (12.59 to 73.08)	14	94
9	0.85 (0.77 to 0.91)	0.82 (0.75 to 0.88)	4.78 (3.30 to 6.93)	0.18 (0.11 to 0.29)	26.68 (13.12 to 54.26)	20	96
10	0.82 (0.76 to 0.86)	0.86 (0.79 to 0.91)	5.95 (3.80 to 9.32)	0.21 (0.16 to 0.28)	28.04 (15.35 to 51.22)	22	97
11	0.72 (0.64 to 0.79)	0.91 (0.85 to 0.95)	8.11 (4.86 to 13.55)	0.31 (0.23 to 0.40)	26.60 (14.73 to 48.02)	21	97
12	0.68 (0.62 to 0.74)	0.93 (0.88 to 0.96)	9.81 (5.94 to 16.18)	0.34 (0.29 to 0.41)	28.57 (16.56 to 49.27)	20	98
13	0.66 (0.57 to 0.74)	0.93 (0.90 to 0.95)	9.08 (6.53 to 12.62)	0.36 (0.29 to 0.46)	24.89 (16.36 to 37.87)	24	97
14	0.53 (0.48 to 0.58)	0.96 (0.92 to 0.98)	12.45 (6.47 to 23.97)	0.49 (0.44 to 0.54)	25.52 (12.66 to 51.45)	12	97
15	0.39 (0.32 to 0.46)	0.98 (0.96 to 0.99)	17.45 (9.83 to 30.98)	0.62 (0.55 to 0.70)	28.01 (14.53 to 54.00)	10	79
16	0.31 (0.19 to 0.47)	0.99 (0.98 to 0.99)	29.13 (10.70 to 79.26)	0.69 (0.56 to 0.86)	41.92 (12.90 to 136.2)	6	41

DOR, diagnostic odds ratio; LR+, positive likelihood ratio; LR–, negative likelihood ratio.
Note: Values in parentheses are 95% confidence intervals.

(i.e. Youden's index), was 10 (*Figure 18*). At this cut point a pooled sensitivity of 0.82 (95% CI 0.76 to 0.86) and specificity of 0.86 (95% CI 0.79 to 0.91) were derived. The positive likelihood ratio associated with the sensitivity and specificity was 5.95 (95% CI 3.80 to 9.32) and the pooled DOR was 28.04 (95% CI 15.35 to 51.22). A summary plot of sensitivity and specificity in ROC space, summarising each study at cut point 10, weighted by study size can be seen in *Figure 19*; additional

plots for the other cut points can be found in Appendix 4. Although all of the studies used the EPDS, used a widely recognised gold standard and were focusing on identifying major or minor depression the I^2 value of 97% was very high. Thus, care should be taken when interpreting the results.

When the studies using the BDI at a single cut point were combined the pooled sensitivity was 0.72 (95% CI 0.54 to 0.85) and specificity was
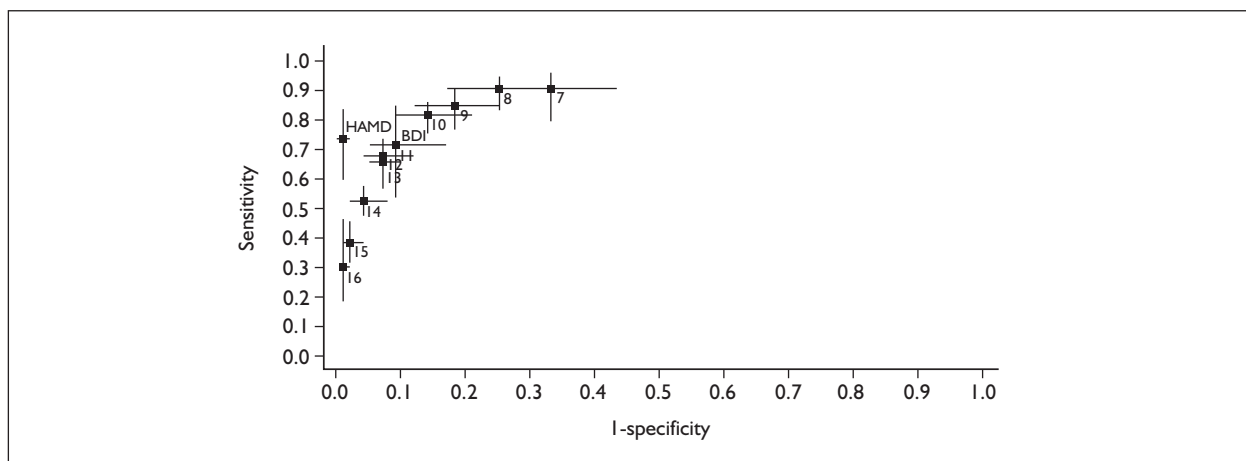
FIGURE 18 *Graphical summary of sensitivity and specificity postnatally at varying cut points for major or minor depression.*

0.91 (95% CI 0.83 to 0.95) for the diagnosis of major or minor depression (*Table 14*). The positive likelihood ratio associated with the sensitivity and specificity was 7.66 (95% CI 3.34 to 17.60) and the pooled DOR was 24.74 (95% CI 5.95 to 102.77). The I^2 value was high at 99%, hence care should be taken when interpreting the results. For the HAMD the pooled sensitivity was 0.74 (95% CI 0.60 to 0.84) and specificity was 0.99 (95% CI 0.98 to 1.00) for the diagnosis of major or minor depression (*Table 14*). The positive likelihood ratio associated with the sensitivity and specificity was 86.83 (95% CI 32.29 to 233.46) and the pooled DOR was 325.93 (95% CI 102.36 to 1037.84). As a single study provided the data to pool for the HAMD (the instrument was administered at different time points on the same women using the same cut point) it is unsurprising that the I^2 value is 0. A summary plot of sensitivity and specificity in ROC space, summarising each study for the BDI and HAMD, weighted by study size can be seen in *Figure 20*.

When the psychometric attributes were pooled across the studies, high levels of between-study heterogeneity were identified (range 41–98%). From *Table 13* we can see that moderate or high levels of between-study heterogeneity were identified across all cut points. Meta-regression was undertaken for cut points 7–15; because of the low number of studies reporting data for cut point 16 we were unable to investigate any potential sources of heterogeneity.

Table 15 shows the results of the univariate analyses. As meta-regression was undertaken within each cut point multiple tests were performed and therefore we must be cautious in interpreting the results. It can be seen that verification bias seemed to be an important factor across cut points 7–13. The DORs for studies undertaking a diagnostic interview for the whole sample (or random sample) were lower than those for studies undertaking diagnostic interviews on a non-random sample. For cut points 14 and 15, blinding seemed to be the most important factor in explaining the heterogeneity in the DOR of the primary studies. The DORs were higher in those studies in which the index test results were interpreted without knowledge of the results of the reference standard than in those studies in which the index test results were interpreted with knowledge of the results of the reference standard. Despite many of these results being statistically significant at the $p \leq 0.05$ level, the I^2 values still remain high.

Table 16 shows the results of the multivariate analyses. As meta-regression was undertaken within each cut point multiple tests were performed and, furthermore, because of the low number of studies included we should be cautious in interpreting the results. Although differences for some variables (verification bias) were statistically significant at some cut points the coefficients were extremely low and I^2 values high, thus indicating that none of the a priori sources of heterogeneity were predictive when entered as covariates in the meta-regression and that there are obviously still some sources of variation that remain unexplained. At cut point 15, when all variables were considered simultaneously in the model the I^2 value reduced from 79% to 0% indicating that these variables explained all of the heterogeneity in the DORs of the primary studies.

Any psychiatric disorder
Any psychiatric disorder was used to classify studies when the disorder under study was not differentiated between major depression or minor
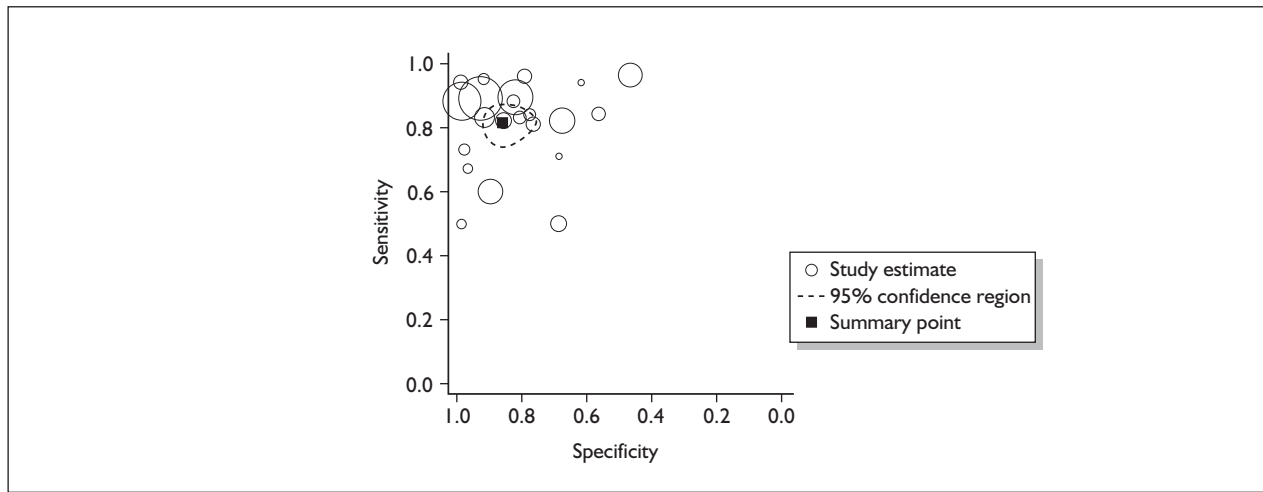
FIGURE 19 *EPDS SROC plot for diagnosis of major or minor depression at cut point 10.*

TABLE 14 *Summary of the sensitivity and specificity of the BDI and HAMD for major or minor depression*

	Sensitivity	**Specificity**	**LR+**	**LR−**	**DOR**	*I²*
BDI	0.72 (0.54 to 0.85)	0.91 (0.83 to 0.95)	7.66 (3.34 to 17.60)	0.31 (0.17 to 0.58)	24.74 (5.95 to 102.8)	99
HAMD	0.74 (0.60 to 0.84)	0.99 (0.98 to 1.00)	86.83 (32.29 to 233.46)	0.27 (0.17 to 0.42)	325.93 (102.4 to 1037.8)	0

BDI, Beck Depression Inventory; DOR, diagnostic odds ratio; HAMD, Hamilton Rating Scale for Depression; LR+, positive likelihood ratio; LR−, negative likelihood ratio.
Note: Values in parentheses are 95% confidence intervals.
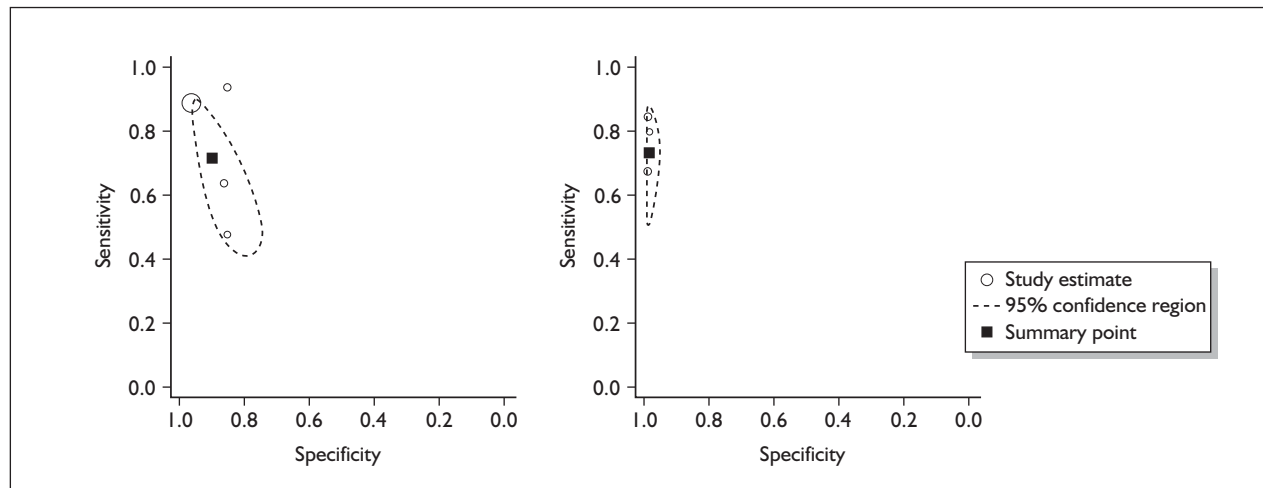
FIGURE 20 *SROC plot for diagnosis of major or minor depression.*

depression or dysthymia, but did not incorporate other types of psychiatric disorder such as anxiety. Eight studies compared identification strategies with a gold standard in the postnatal period for the diagnosis of what we have labelled 'any psychiatric disorder'. *Table 17* summarises the types of disorder that fall into this category. Two identification strategies were used: EPDS (*n* = 8) and BDI-II (*n* = 1). Insufficient data were available to permit

pooling for the BDI-II. The EPDS had sufficient data to pool at a variety of cut points (9–14). A summary of the sensitivity and specificity of all studies is summarised in *Figure 21*.

When the studies were combined the pooled sensitivities ranged from 0.38 (95% CI 0.28 to 0.48) to 0.86 (95% CI 0.75 to 0.92) and specificities ranged from 0.85 (95% CI 0.78 to 0.90) to 0.99

TABLE 15 *Univariate meta-regression analyses*

Cut point	Predictor	DOR_1	DOR_2	Beta-coefficient (95% CI)	*p*-value	I^2
7	Prevalence	–	–	2.25 (–8.99 to 13.49)	0.66	75
	Verification bias	6.07	70.15	0.08 (0.02 to 0.45)	0.01	49
	Blinding	22.77	8.06	2.21 (0.22 to 21.83)	0.45	75
	Timing	10.77	13.23	0.71 (0.06 to 8.91)	0.76	74
8	Prevalence	–	–	–2.99 (–14.29 to 8.31)	0.58	94
	Verification bias	8.25	85.92	0.12 (0.03 to 0.50)	0.008	80
	Blinding	26.94	26.93	1.24 (0.11 to 14.61)	0.85	94
	Timing	32.51	13.64	1.55 (0.12 to 19.12)	0.71	94
9	Prevalence	–	–	–1.54 (–10.02 to 6.94)	0.71	96
	Verification bias	10.25	146.70	0.09 (0.03 to 0.32)	0.001	86
	Blinding	36.30	22.71	2.03 (0.24 to 17.13)	0.50	96
	Timing	28.88	18.30	1.19 (0.21 to 6.68)	0.83	96
10	Prevalence	–	–	–2.22 (–8.04 to 3.59)	0.44	96
	Verification bias	14.99	70.00	0.24 (0.07 to 0.76)	0.02	92
	Blinding	37.45	29.15	1.36 (0.23 to 7.99)	0.72	97
	Timing	38.51	21.15	1.50 (0.32 to 7.05)	0.59	97
11	Prevalence	–	–	0.42 (–4.68 to 5.52)	0.87	94
	Verification bias	15.89	47.10	0.33 (0.10 to 1.08)	0.07	90
	Blinding	36.40	19.06	1.67 (0.31 to 8.91)	0.53	97
	Timing	19.94	27.68	0.64 (0.14 to 2.93)	0.54	98
12	Prevalence	–	–	–0.02 (–5.18 to 5.14)	0.99	96
	Verification bias	22.86	39.25	0.53 (0.15 to 1.93)	0.32	94
	Blinding	33.87	22.74	1.27 (0.27 to 5.84)	0.75	98
	Timing	29.67	26.59	1.00 (0.24 to 4.24)	0.99	98
13	Prevalence	–	–	–1.87 (–5.97 to 2.23)	0.35	92
	Verification bias	17.05	46.43	0.35 (0.15 to 0.81)	0.02	85
	Blinding	38.03	22.93	1.33 (0.27 to 6.47)	0.72	97
	Timing	20.14	29.82	0.56 (0.13 to 2.44)	0.42	97
14	Prevalence	–	–	–2.98 (–13.33 to 7.36)	0.54	97
	Verification bias	14.18	43.59	0.32 (0.07 to 1.51)	0.13	87
	Blinding	66.49	13.85	4.97 (1.18 to 20.99)	0.03	83
	Timing	27.56	22.76	1.22 (0.21 to 6.99)	0.81	93
15	Prevalence	–	–		0.29	74
	Verification bias	34.16	37.72	0.83 (0.15 to 4.59)	0.81	79
	Blinding	66.70	23.75	3.16 (0.88 to 11.37)	0.07	53
	Timing	44.09	29.36	1.58 (0.33 to 7.70)	0.52	71

CI, confidence interval; DOR, diagnostic odds ratio.

(95% CI 0.97 to 1.00) for the diagnosis of any psychiatric disorder at a range of cut points using the EPDS (*Table 18*). The optimal cut point, in terms of the trade-offs between sensitivity and specificity (i.e. Youden's index), was 9 (*Figure 22*). At this cut point a pooled sensitivity of 0.86 (95% CI 0.75 to 0.92) and specificity of 0.87 (95% CI 0.73 to 0.94) were derived. The positive likelihood ratio associated with the sensitivity and specificity was 6.54 (95% CI 3.19 to 13.43) and the pooled DOR was 39.46 (95% CI 18.98 to 82.06). A summary plot of sensitivity and specificity in ROC space,

TABLE 16 *Multivariate meta-regression analyses*

Cut point	Predictor	Beta-coefficient (95% CI)	*p*-value	I^2
7	Prevalence	3.24 (−13.12 to 19.59)	0.63	47
	Verification bias	0.05 (0.002 to 1.27)	0.06	
	Blinding	0.42 (0.01 to 22.97)	0.60	
	Timing	0.89 (0.08 to 10.37)	0.91	
8	Prevalence	−1.32 (−14.43 to 11.79)	0.83	84
	Verification bias	0.09 (0.01 to 1.47)	0.08	
	Blinding	0.39 (0.02 to 8.69)	0.51	
	Timing	1.37 (0.10 to 18.39)	0.79	
9	Prevalence	−0.90 (−7.47 to 5.67)	0.78	87
	Verification bias	0.06 (0.01 to 0.32)	0.003	
	Blinding	0.41 (0.06 to 2.62)	0.32	
	Timing	0.88 (0.26 to 3.00)	0.82	
10	Prevalence	−1.06 (−6.71 to 4.59)	0.70	92
	Verification bias	0.18 (0.04 to 0.92)	0.04	
	Blinding	0.52 (0.08 to 3.19)	0.45	
	Timing	1.28 (0.31 to 5.22)	0.71	
11	Prevalence	−0.49 (−5.66 to 4.68)	0.84	88
	Verification bias	0.39 (0.07 to 2.17)	0.27	
	Blinding	1.72 (0.24 to 12.34)	0.57	
	Timing	0.38 (0.09 to 1.62)	0.17	
12	Prevalence	0.02 (−6.14 to 6.18)	0.995	95
	Verification bias	0.58 (0.09 to 3.61)	0.54	
	Blinding	1.21 (0.15 to 9.80)	0.85	
	Timing	0.81 (0.15 to 4.31)	0.80	
13	Prevalence	−1.40 (−5.49 to 2.70)	0.48	84
	Verification bias	0.44 (0.14 to 1.39)	0.15	
	Blinding	1.57 (0.35 to 6.93)	0.53	
	Timing	0.45 (0.13 to 1.53)	0.19	
14	Prevalence	−4.47 (−14.22 to 5.28)	0.31	85
	Verification bias	1.31 (0.07 to 23.07)	0.83	
	Blinding	7.09 (0.28 to 181.88)	0.20	
	Timing	0.56 (0.07 to 4.35)	0.52	
15	Prevalence	−10.14 (−17.75 to −2.52)	0.02	0
	Verification bias	5.03 (0.94 to 26.97)	0.06	
	Blinding	27.18 (4.18 to 176.66)	0.01	
	Timing	0.12 (0.02 to 0.81)	0.04	

CI, confidence interval

summarising each study at cut point 9, weighted by study size can be seen in *Figure 23*; additional plots for the other cut points can be found in Appendix 4. When the psychometric attributes were pooled across the studies, a variety of levels of between-study heterogeneity were identified (range 0–60%).

From *Table 18* we can see that, apart from cut point 10, the I^2 values were relatively low, indicating low levels of between-study heterogeneity. Because of the low numbers of studies reporting data at each cut point care should be taken when interpreting the results.

TABLE 17 *'Any psychiatric disorder' classifications*

Study	Types of disorder
Abiodun, 2006[126]	Postnatal depression
Aslan *et al.*, 1997[143]	Depression
Benvenuti *et al.*, 1999[134]	Major depression, anxiety and mood disorders
Carpiniello *et al.*, 1997[135]	Clinically depressed
Ghubash *et al.*, 1997[156]	Catego definition of depression
Teng *et al.*, 2005[155]	Major depressive disorder, depressive disorder not otherwise stated and dysthymic disorder
Uwakwe and Okonkwo, 2003[127]	Affective morbidity
Zelkowitz and Milet, 1995[37]	Any depressive disorder
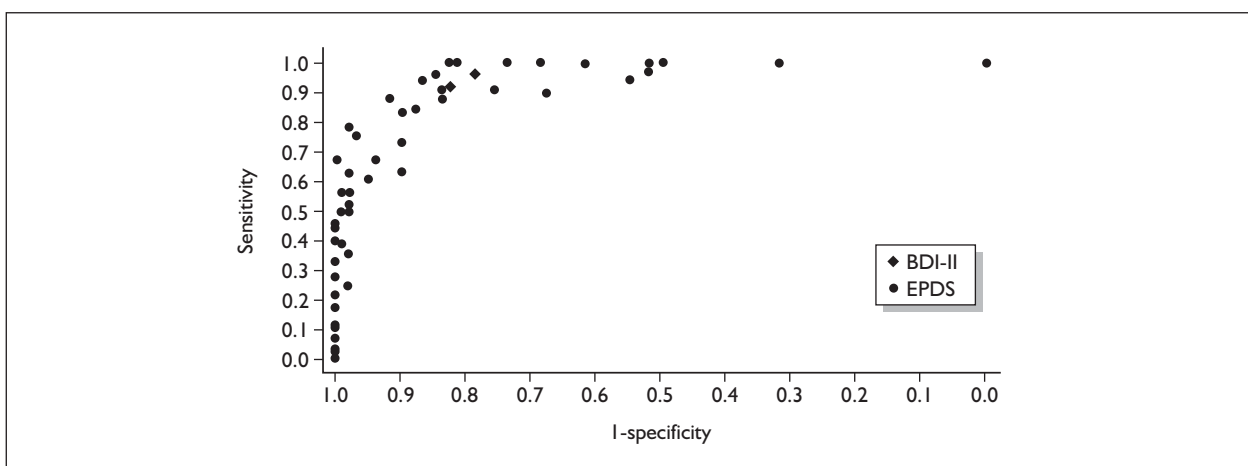
FIGURE 21 *Summary of identification strategies used postnatally at varying cut points to diagnose any psychiatric disorder.*

TABLE 18 *Summary of the sensitivity and specificity of the EPDS postnatally at varying cut points for any psychiatric disorder*

Cut point	Sensitivity	Specificity	LR+	LR–	DOR	*n*	*I*²
9	0.86 (0.75 to 0.92)	0.87 (0.73 to 0.94)	6.54 (3.19 to 13.43)	0.17 (0.10 to 0.28)	39.46 (18.98 to 82.06)	4	44
10	0.81 (0.64 to 0.91)	0.85 (0.78 to 0.90)	5.34 (3.84 to 7.43)	0.23 (0.12 to 0.44)	23.35 (11.88 to 45.91)	4	60
11	0.70 (0.53 to 0.83)	0.93 (0.86 to 0.97)	10.32 (5.23 to 20.36)	0.32 (0.20 to 0.52)	32.13 (14.84 to 69.60)	4	0
12	0.57 (0.48 to 0.65)	0.97 (0.96 to 0.98)	21.08 (12.73 to 34.92)	0.44 (0.36 to 0.54)	47.62 (26.70 to 84.94)	5	0
13	0.66 (0.47 to 0.80)	0.96 (0.91 to 0.98)	17.35 (8.25 to 36.51)	0.36 (0.22 to 0.57)	48.52 (25.15 to 93.63)	7	0
14	0.38 (0.28 to 0.48)	0.99 (0.97 to 1.00)	33.38 (11.87 to 93.88)	0.63 (0.54 to 0.74)	52.86 (17.17 to 162.77)	4	25

DOR, diagnostic odds ratio; LR+, positive likelihood ratio; LR–, negative likelihood ratio.
Note: Values in parentheses are 95% confidence intervals.

51

Other classifications

Other classifications incorporated studies that considered depression alongside a mixed group of related psychiatric disorders, such as anxiety, and did not report psychometric properties for depression alone. Five studies were identified that used identification strategies to identify other types of disorders.[34,35,115,121,130] Pooling was not undertaken as each study was classifying different disorders (*Table 19*); however, a summary of the sensitivities and specificities are shown in *Figure 24*. The earliest study[115] focused on identifying all RDC diagnoses using the GHQ. Women were examined twice during the postnatal period, at 5 days and 1 month postnatally. Sensitivity and specificity values were reported for a range of cut points for both time points. Sensitivity ranged from 0.11 (specificity 0.96) to 1.00 (specificity 0.08) at 5 days postnatally and from 0.25 (specificity 0.94) to 1.00 (specificity 0.05) at 1 month postnatally. The authors set an optimal cut point of 8, giving a sensitivity of 0.28 (specificity 0.79) at 5 days postnatally and a sensitivity of 0.50 (specificity 0.84) at 1 month postnatally.

The second study[34] focused on identifying women with depressed mood alone, neurotic or psychotic depression and other diagnoses using the EPDS at 3 months postnatally. Sensitivity and specificity values were reported for two cut points using four
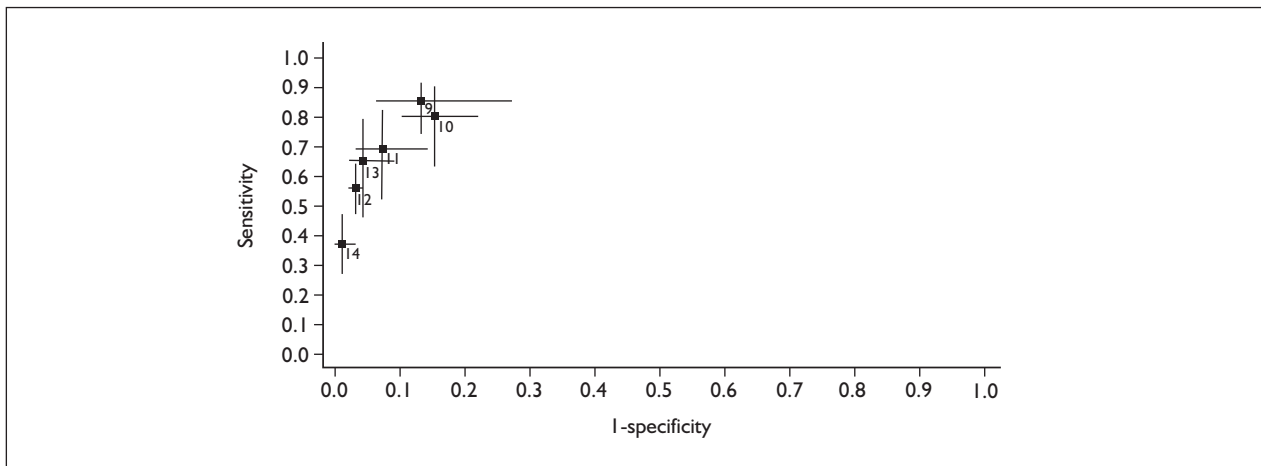
FIGURE 22 *Graphical summary of pooled sensitivity and specificity of the EPDS postnatally at varying cut points to diagnose any psychiatric disorder.*
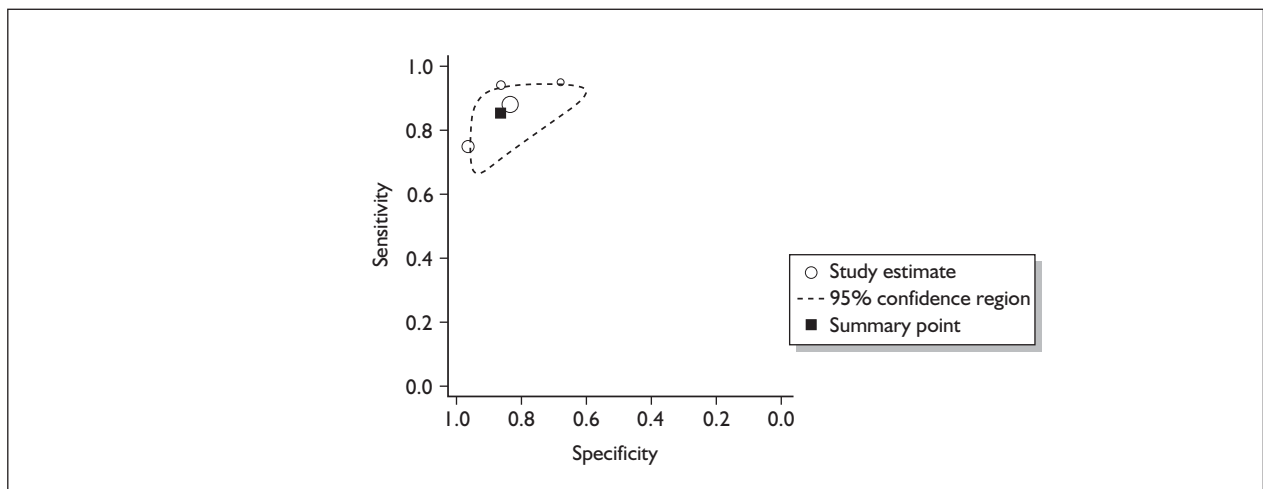
FIGURE 23 *EPDS sROC plot for diagnosis of any psychiatric disorder at cut point 9.*

TABLE 19 *Other classifications*

Study	Types of disorders
Kitamura *et al.*, 1994[115]	All RDC diagnoses
Leverton and Elliott, 2000[34]	Depressed mood alone, neurotic or psychotic depression and other diagnoses
Matthey *et al.*, 2001[35]	Depressive and anxiety disorders
Milgrom *et al.*, 2005[121]	Major depressive disorder, depressive disorder NOS, adjustment disorder with depression, mixed anxiety depressive disorder, bipolar disorder, psychotic disorder and dysthymic disorder
Navarro *et al.*, 2007[130]	Mood disorders, anxiety disorders, adjustment disorders and other diagnoses
NOS, not otherwise stated; RDC, Research Diagnostic Criteria.	
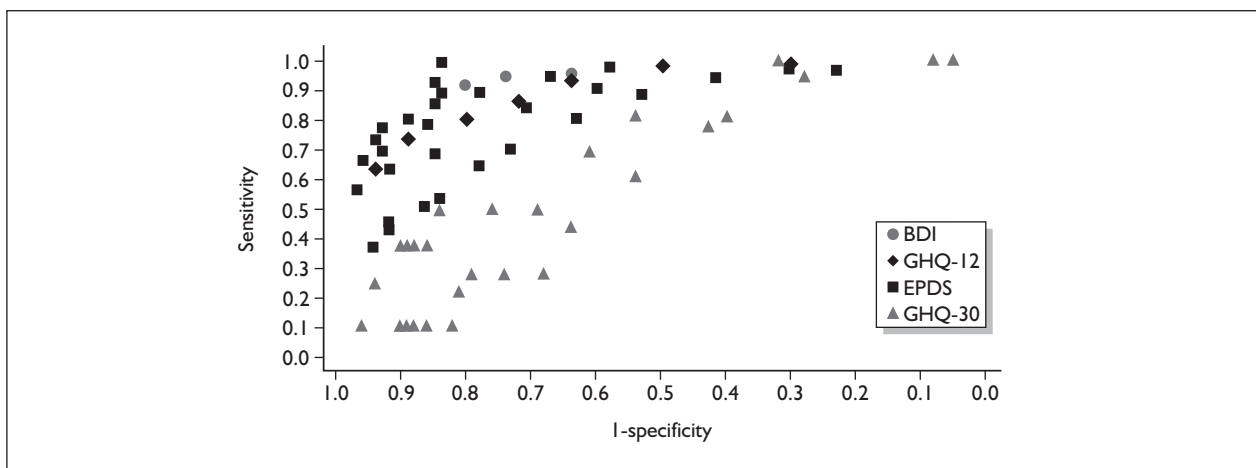
FIGURE 24 *Summary of identification strategies used postnatally at varying cut points to diagnose other classifications.*

versions of the gold standard (Catego any time in the month, Bedford College any time in the month, Catego last 2 weeks of the month, Bedford College last 2 weeks of the month). Sensitivity ranged from 0.64 (specificity 0.92) to 1.00 (specificity 0.84).

The third study[35] focused on depressive and anxiety disorders, aiming to validate the EPDS for use among women and their partners; the results summarised here relate only to the performance of the EPDS in women. Sensitivity and specificity values were reported for various cut points. Sensitivity ranged from 0.38 (specificity 0.95) to 0.97 (specificity 0.23). The authors found an optimal cut point of 8, giving a sensitivity of 0.70 and a specificity of 0.73.

The next study[121] focused on identifying women with major depressive disorder, depressive disorder not otherwise specified, adjustment disorder with

depression, mixed anxiety depressive disorder, bipolar disorder, psychotic disorder and dysthymic disorder using the EPDS and BDI at 4 months postnatally. Sensitivity and specificity values were reported at three cut points for both instruments. Sensitivity ranged from 0.92 (specificity 0.80) to 0.96 (specificity 0.64) for the BDI and from 0.79 (specificity 0.86) to 0.91 (specificity 0.60) for the EPDS. In this population the BDI had higher diagnostic efficiency than the EPDS.

In the final study[130] the EPDS and GHQ-12 were used to identify women with mood disorders, anxiety disorders, adjustment disorders and other diagnoses at 6 weeks postnatally. Sensitivity and specificity were reported at a range of cut points for both instruments. Sensitivity ranged from 0.64 (specificity 0.94) to 0.99 (specificity 0.30) for the GHQ-12 and from 0.57 (specificity 0.97) to 0.98 (specificity 0.58) for the EPDS. The authors found an optimal cut point of 5 for the GHQ-12, giving

a sensitivity of 0.81 and a specificity of 0.80, and a cut point of 10 for the EPDS, giving a sensitivity of 0.86 and a specificity of 0.85.

Discussion

In total, 14 identification strategies have been validated among women during pregnancy or the postnatal period (up to 1 year). Identification strategies included PND-specific measures and also generic depression identification strategies (specific: EPDS, PDSS, PRQ, PI; generic: BDI, GHQ, HADS, HSCL, HAMD, Zung's SDS, SCL-90-R, Raskin, MADRS; other: EPDS–GHQ double test). By far the most frequently used identification strategy was the EPDS. Quality assessment was undertaken using QUADAS and there was variability in the results across the individual items. It is interesting to note that none of the studies fulfilled all of the quality criteria. Studies demonstrated high quality in five out of the eight questions focusing on bias (questions 3, 5, 6, 7, and 14) and the three questions relating to reporting quality (questions 8, 9 and 13). The poorest quality rating was associated with question 10, regarding whether the index test results were interpreted without knowledge of the reference standard results. This item is important as interpretation of the results of the index test may be influenced by knowledge of the results of the reference standard leading to inflated estimates of diagnostic accuracy. Two other poorly reported items were selection criteria and disease progression bias.

There were sufficient data for postnatal studies across a large number of cut points of the EPDS to be able to combine results and produce pooled summary estimates of sensitivity and specificity. However, there were insufficient data at each cut point for most other identification strategies to be able to pool data. For major depression only, the pooled estimates of sensitivity and specificity for the EPDS ranged from 0.60 to 0.96 and from 0.45 to 0.97 respectively. For any psychiatric disorder, the pooled estimates of sensitivity and specificity for the EPDS ranged from 0.38 to 0.86 and from 0.85 to 0.99 respectively. Finally, for major or minor depression, the pooled estimates of sensitivity and specificity for the EPDS ranged from 0.31 to 0.91 and from 0.67 to 0.99 respectively. In addition, for major or minor depression there were sufficient data to pool the BDI and HAMD at a single cut point. Results from this analysis highlighted that generic identification strategies may be less sensitive than the EPDS, but more

specific. Caution should be taken when interpreting these results as there were only four studies included for the generic identification strategies and also the cut points used may not be the best cut points to use to identify women with PND.

When the psychometric attributes were pooled across the studies, high levels of between-study heterogeneity were identified in most analyses (major depression: 0–93%; major or minor depression: 41–98%). Unfortunately, none of the a priori sources of heterogeneity were predictive in a meta-regression analysis, and high levels of between-study heterogeneity remained in the model. There was some suggestive evidence that the timing of administration of the EPDS (within 6 weeks postnatally or not) may be an important factor in influencing diagnostic performance. Two other variables, verification bias and blinding, also demonstrated some potential effects on diagnostic performance, hence any future validation studies should be undertaken using methods to avoid such biases. Further research would be informative to identify key sources of heterogeneity and specifically whether different items need to be considered when pooling psychometric instruments in diagnostic accuracy studies.

There are limitations to the review. In many cases multiple data points were presented for the studies evaluated. Multiple data points arose for many reasons: more than one identification strategy was used; more than one cut point was presented; the identification strategy was repeatedly administered; two versions of the identification strategy were used; different classifications of depression were recorded; more than one reference standard was used. Studies were pooled at individual cut points in an attempt to overcome the fact that multiple data points were presented. The drawbacks of pooling at individual cut points were multiple testing and fewer studies included, thus reductions in power. Furthermore, because of the low number of studies included at some cut points, meta-regression could not be undertaken and potential sources of heterogeneity could not be explored.

A unique issue that arises when undertaking meta-analyses of diagnostic accuracy studies is variations in the cut point chosen to indicate a positive test. The higher the cut point value chosen, the higher the specificity and the lower the sensitivity will be. Threshold effects create a potential source of heterogeneity and to reduce this we pooled studies within individual cut points. However, despite pooling within cut points, threshold effects

were still apparent. Methods to simultaneously model thresholds within thresholds would be useful. Finally, difficulties arose in defining the appropriate cut points to be used in practice. The relevant cut point depends on the viewpoint taken: statistical, clinical or economic. From a statistical viewpoint it would seem sensible to find the optimal values based on the trade-off between sensitivity and specificity (e.g. Youden's index). From a clinical perspective it may be more important, in this situation of identifying PND, for the identification strategy to be more sensitive than specific. Nevertheless, maximising sensitivity estimates will lead to subsequent reductions in the specificity estimates resulting in more false positives being identified. Increasing the number of false positives will increase costs and resource use within the NHS.

The identification strategies reviewed here appear to be able to identify PND in women during pregnancy and the postnatal period with a degree of accuracy that is similar, if not slightly better, to that for depression in the general population. In an evaluation of case-finding instruments for identifying patients with major depression or dysthymia in primary care,[83] 16 instruments were assessed in 38 studies and the overall sensitivity was 0.79 (95% CI 0.74 to 0.83) and overall specificity was 0.75 (95% CI 0.70 to 0.81). Equivalent estimates from this review for the EPDS resulted in overall estimates of sensitivity of 0.86 (95% CI 0.81 to 0.89) and of specificity of 0.87 (95% CI 0.80 to 0.92). In summary, the EPDS is the most frequently reported identification strategy and its diagnostic performance seems reasonably good.

Reflection on current policy and practice within the UK

As outlined in Chapter 1 current NICE guidance recommends the use of two questions to identify possible depression and a third question if the women answers 'yes' to either of the initial questions [(1) 'During the past month, have you often been bothered by feeling down, depressed or hopeless?'; (2) 'During the past month, have you often been bothered by little interest or pleasure in doing things?'; and (3) 'Is this something you feel you need or want help with?'). No studies were identified in this review that validated this recommended method to identify women with PND. Studies undertaken to validate the Whooley questions plus the help question in a general depressed population (including both men and women) found that the sensitivity was 0.96 (95% CI 0.86 to 0.99) and the specificity was 0.89 (95% CI 0.87 to 0.91).[39] The diagnostic performance of the three questions in a depressed population was better than the performance of the EPDS in a postnatal population; however, until further research is undertaken we cannot be confident that these results will be maintained when the three questions are used in a postnatal population. Within the NICE guidance, the psychometric properties of the EPDS postnatally were based on eight validation studies identified from a literature search and a single systematic review (which included eight validation studies). Within this systematic review and meta-analysis we identified 48 validation studies of the EPDS post-natally.

Chapter 6

Acceptability to women and health professionals of methods to identify postnatal depression: systematic review 2

In the absence of an existing review of PND identification strategies, we comprehensively and systematically synthesised the evidence from qualitative and quantitative research that addressed the question, 'Are PND identification strategies acceptable to women and health professionals?'

Methods

Inclusion criteria

We included studies that assessed the acceptability of PND identification to women or health professionals in the prenatal or postnatal period (up to 1 year). The acceptability of these identification strategies was assessed in studies using two approaches: (1) asking women or health professionals about their views on these strategies using qualitative or quantitative methods; or (2) recording refusal or non-participation rates on application of a standardised questionnaire such as the EPDS. For the purpose of our review, qualitative research was defined as those studies that collected data about women's or health professionals' views using specific qualitative techniques such as unstructured interviews, semistructured interviews or focus groups. Quantitative research used survey methods to assess women's or health professionals' views on the identification strategies, or to record refusal or non-participation rates for the completion of a standardised questionnaire. Studies were also eligible for inclusion if both qualitative and quantitative approaches were used, that is, they used mixed methods. No studies were excluded on the grounds of quality, which is in line with the pragmatic choice of other reviewers.[157,158]

Data extraction

Two reviewers independently assessed the titles and/or abstracts of the citations from the electronic searches. For potentially eligible studies we retrieved the papers and used an electronic proforma to record study eligibility and to extract data. Both reviewers independently assessed these studies for inclusion and resolved any disagreement through discussion.

Quality assessment

There are various strategies available for qualitative research to protect against bias and enhance the reliability and external validity of findings, which are summarised in checklists.[159–161] In our review these checklists were used to inform data extraction from eligible qualitative studies, which included study design, method of sampling and description of the sample, timing of data collection, and type of analysis. There are also accepted criteria for appraising the quality of quantitative studies such as surveys.[162] We chose criteria to extract data about the surveys, to complement the data extracted from the qualitative studies. This included how the sample for the survey was obtained, a description of the sample, timing of data collection, and type of analysis. For both qualitative and quantitative eligible studies, data were extracted about the identification strategy. This included information regarding the type of instrument used, such as the EPDS or GHQ; the timing of the strategy, such as pre- or postnatal; the setting where the strategy was administered, such as an antenatal clinic or women's homes; and the mode of data collection, such as interview or self-completed questionnaire. Each of the two reviewers independently performed the extraction of data for a sample of studies and when there was uncertainty this was resolved through discussion. Authors were contacted when necessary for further information.

Data synthesis

The synthesis of qualitative data in systematic reviews is an area of ongoing methodological development.[163–165] A recent peer-reviewed publication described two alternative methods for synthesising evidence from qualitative studies in systematic reviews: 'textual narrative synthesis' and 'thematic synthesis'.[166] The textual narrative approach involves grouping studies together into

subgroups; writing a commentary on key aspects of studies in relation to the subgroup within which they were included; and then conducting subgroup synthesis. Thematic synthesis involves each reviewer independently identifying themes that arise in relation to research questions; comparing themes produced by each reviewer; clustering themes together under each research question; and then agreeing on synthesis of evidence from studies for each theme.

To synthesise the qualitative evidence from studies included in our review we adapted the textual narrative approach as this method has been found to be particularly successful in synthesising different types of research evidence (e.g. qualitative and quantitative).[166] This approach required the research team to develop a topic list from our existing knowledge of the research literature that identified factors likely to affect women's or health professionals' views on the acceptability of PND identification strategies. These topics were then used to define a number of subgroups. Studies belonging to each subgroup were identified by one reviewer and written up as a commentary. Study characteristics and quality, for both the qualitative and quantitative studies, were presented in tables separately for those studies that assessed women's views of identification strategies and those that assessed health professionals' views. Narrative synthesis of the evidence for studies in each subgroup was then conducted to help draw conclusions across the qualitative and quantitative studies. For the quantitative data the percentages of non-responses for the completion of a questionnaire such as the EPDS were also reported as an overall indication of the acceptability of an identification strategy.

The value of including data from different types of studies in systematic reviews is increasingly recognised and an approach has been described to combine qualitative and quantitative research.[159] For our review we collected data from qualitative studies that examined in detail a small sample of women's or health professionals' views about the acceptability of identification strategies and from surveys that quantify the acceptability of these strategies for a sample of several hundred women or health professionals. Therefore, having synthesised the evidence from qualitative and quantitative research separately we attempted, when possible, to integrate the findings from the qualitative synthesis of the textual data with the findings from the surveys. We did this by exploring whether or not the themes discussed in qualitative

research studies were included in the surveys and how the presence or absence of these themes in the surveys affected the estimates of acceptability of identification strategies.

Results

In total, 16 studies were eligible and are presented in *Tables 20–25*, which include information on study characteristics and results. Three studies assessed both women's and health professionals' views and so are repeated in the respective tables. Studies are reported according to the timing of the administration of the strategy, either postnatally or prenatally. Studies that administered the identification strategy both before and after the women gave birth were called perinatal studies. None of the included studies assessed non-response rates for the completion of a questionnaire as an indication of the acceptability of a strategy.

Topics that were identified from the research literature as likely to affect views on identification strategies were used to define a number of subgroups. These were:

- method of administration of the strategy (including expectation/awareness that the strategy was to be implemented; timing, setting and mode of administration; feedback of results)
- difficulties in answering questions, such as being a sensitive topic, or fear of disclosure/ being honest
- interpersonal relationships between women and health professionals
- cultural or ethnic differences
- training issues (particularly when exploring health professionals' views)
- overall acceptability of the identification strategy.

Studies belonging to each subgroup were identified and synthesised in a narrative fashion according to the timing of the identification strategy (postnatal, prenatal, or perinatal) and whether the studies reflected the views of women or health professionals.

Women's views of postnatal identification strategies

The largest group of studies (7 out of 16) explored the views of women on postnatal identification strategies (*Table 20*). All seven studies used

TABLE 20 *Women's views of postnatal identification strategies*

Study, country, grey literature?	Aim	Instruments, version	Design, study setting	Method of sampling, sample	Administration of strategy: timing, setting, mode	Timing of data collection, type of analysis	Conclusions
Matthey et al., 1997[170] Australia No	To assess suitability of DIS as a criterion measure of depression	EPDS, GHQ-30, Faces, DIS Vietnamese, Arabic or English	Semistructured interviews Antenatal clinics	Convenience Five women from three groups (Arabic, Vietnamese, Anglo-Celtic); mean age of 28 years	Between 6 weeks and 6 months postnatally Home Self-completed questionnaire	Between 8 and 10 months postnatally Conceptualised in terms of themes	Anglo-Celtic clients did not find any questions culturally inappropriate, but the Vietnamese and Arabic women did
Poole et al., 2006[168] England No	To explore women's views about PND	EPDS English	Semistructured interviews Primary care trust	Purposive 15 women; mean age of 30 years; EPDS scores ranged from 1 to 23	Within a year postnatally Home Self-completed questionnaire	Within a year postnatally Interpretative phenomenological analysis	Women were generally positive about the identification strategy, but had concerns about the instrument itself and lack of information
Shakespeare et al., 2003[167] England No	To explore acceptability to women of identifying PND by health visitors	EPDS English	In-depth interviews General practice	Purposive 39 women; mean age of 34 years; 37 of 39 were white	8 weeks and 8 months postnatally Clinic or home Self-completed questionnaire	Around 15 months postnatally Constant comparative method	21 of 39 women (54%) found the identification strategy unacceptable and had problems with the venue, personal intrusion, and stigma
Cubison and Munro, 2005[169] England Yes	To explore women's experiences of being identified for PND	EPDS 6-item English	Semistructured interviews National Childbirth Trust groups and community mental health service	Convenience and purposive 15 women and four severely postnatally depressed women, all white British	Between 4 and 6 weeks postnatally Unclear Self-complete or in discussion with the health visitor	Unclear Constant comparative method	Women valued the health visitors' attempts to detect PND, but were unhappy with its administration

TABLE 20 *Women's views of postnatal identification strategies (continued)*

Study, country, grey literature?	Aim	Instruments, version	Design, study setting	Method of sampling, sample	Administration of strategy: timing, setting, mode	Timing of data collection, type of analysis	Conclusions
Clarke, 2003[124] Canada Yes	To explore women's opinions of the cultural relevance and sensitivity of instruments	EPDS, PDSS, BDI-II English	Structured interviews Community health centres	Convenience 104 mothers; mean age of 24 years; First Nations and Métis women	Between 1 and 12 months postnatally Community health centres Self-completed questionnaire	After completing questionnaires and interview Coding developed post hoc	Most women thought that the questionnaires were appropriate and culturally sensitive and that they did not miss anything
Gemmill et al., 2006[171] Australia No	To measure acceptability of the EPDS in women	EPDS English	Postal or telephone survey, face-to-face interviews Community	Convenience 479 women (mean age 30 years); infants' average age was 17 weeks (21 of these women surveyed by telephone and 26 women interviewed face-to-face)	Around 4 months postnatally Maternal and child health centres Questionnaire completed with nurse	Survey data collected on average 58 weeks after postnatal completion of EPDS Content analysis and survey responses as percentages	EPDS has good acceptability with 97% of women finding the identification strategy desirable and 81% finding the process 'comfortable' to 'very comfortable'
Werrett and Clifford, 2005[117] England No	To ascertain women's views of the usefulness of the EPDS	EPDS English and Punjabi	Interview Clinics at health-care trusts	Convenience 23 women (mean age 29 years); 20 women were married and three were single parents	5–8 weeks postnatally (English), 10–14 weeks postnatally (Punjabi) Unclear Self-completed questionnaire	1 week after the completion of the questionnaire at 5–8 weeks Content analysis and coding developed post hoc	The EPDS was acceptable to the majority of mothers, but may be more applicable to mothers for whom Punjabi is their first language

BDI, Beck Depression Inventory; DIS, Diagnostic Interview Schedule; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; PDSS, Postpartum Depression Screening Scale; PND, postnatal depression.

the EPDS and two studies also included other instruments used to identify PND. Interviews were used as the method of collecting data in all seven studies and they were mainly conducted in primary care. Most women were invited to take part from a readily accessible population of women, that is, they were recruited from a convenience sample. The mean age of women varied from 24 to 34 years and, although most women appeared to be white English-speaking, two studies specifically targeted other ethnic groups. The timing of administration of the questionnaire varied from 1 to 12 months postnatally and the setting was in women's homes or community centres. For six studies the identification strategy was administered as a self-completed questionnaire. The collection of data about women's views ranged from 1 month to around 15 months postnatally, and various analyses were employed.

Qualitative synthesis
Method of administration of postnatal depression identification strategies

For two[167,168] of the seven studies there was discussion about whether women knew that they were going to be assessed for PND. Shakespeare and colleagues[167] found that some, but not all, women had been informed about why they were being asked to complete the questionnaire. If they felt poorly prepared they were anxious about the consequences and reluctant to answer the questions honestly. Poole and colleagues[168] found that most women expected to be assessed for PND and accepted this as part of routine care, but two women who were not expecting this were apprehensive.

Only one study[168] addressed the timing of administration. Poole and colleagues[168] observed that half of the mothers in their sample considered the timing of the first completion of the questionnaire at around 8 weeks postnatally to be appropriate. Mothers highlighted how some negative thoughts and feelings were to be expected during the first few weeks of new motherhood, and so there needed to be a period of 'adaptation' before the use of postnatal identification strategies. In contrast, mothers who had experience of PND felt that depression should have been detected sooner via an early identification programme.

The context and setting in which the routine administration of standardised questionnaires occurred was felt to be important. Of the two studies that explored the setting of the identification strategy, Poole and colleagues[168]

found that all 15 mothers interviewed thought that the completion of the questionnaire should take place at home. Those mothers who completed the questionnaire in a clinic thought that there were too many distractions and were uncomfortable about discussing responses. Similarly, Shakespeare and colleagues[167] found that the 34 women who completed the questionnaire in a clinic found this unacceptable for reasons such as lack of time and privacy, whereas the three women who completed the questionnaire at home found the experience more acceptable.

Three studies addressed the mode of collecting data.[167–169] Shakespeare and colleagues[167] reported that women who had found completing the questionnaire acceptable had little to say about it, but women who had more negative views suggested that they would have preferred open questions and the opportunity to talk rather than be asked to complete a standardised questionnaire such as the EPDS. Furthermore, Cubison and Munro[169] found from their interviews of 15 women that most were critical of a standardised questionnaire with multiple choice tick boxes. They saw it as 'impersonal', 'crude', 'brutal', 'blunt' and 'clumsy', and suggested the option for open questions or to talk about their feelings. In addition, Poole and colleagues[168] found mothers were critical of the lack of dialogue that could result from using a pen and paper assessment.

Two studies[167,168] highlighted the issue of 'feeding back' to women the results of the EPDS. Poole and colleagues[168] found that not all women were aware of their EPDS score, and two mothers who were not provided feedback that they felt that this was unsatisfactory.[168] A minority of mothers who were informed that the results were 'high' reported being 'relieved' as it helped them to understand their difficulties and enabled them to get help. In contrast, other mothers were concerned about high scores and the consequences of being identified as depressed. Shakespeare and colleagues[167] found that the attitude of the health visitor and feedback of results were important to women.[167] About half of the women felt listened to and found it helpful to talk freely with the health visitor. The one-third who had little feedback felt dissatisfied because the way that they were feeling had not been adequately addressed or they sensed that the health visitor was short of time or uninterested.

Difficulties in answering questions
Three[168–170] of the seven studies made criticisms about answering specific questions used in the

instruments to identify PND. In particular, Matthey *et al.*[170] discussed in detail the difficulty that a sample of five Anglo-Celtic, Vietnamese and Arabic women had with answering specific questions on the GHQ-30, the EPDS, the Faces scale and the DIS – a structured psychiatric interview designed for administration by lay personnel to provide DSM-III diagnoses. The Anglo-Celtic women considered that all of the questions on the four instruments were culturally appropriate and were likely to result in a woman responding openly. However, concerns were raised about the cultural meaning of depression for some respondents. For example, among Vietnamese women it was felt that the GHQ asked certain questions that were either inappropriate or would not elicit true feelings from depressed women because to admit to these feelings would bring unbearable shame. These questions were Q12 ('felt that you are playing a useful part in things'), Q13 ('felt capable about making decisions about things'), Q18 ('been taking things hard') and Q24 ('been thinking of yourself as a worthwhile person'). For the EPDS, Q10 ('the thought of harming myself has occurred to me') was also considered an inappropriate item because of the extent of shame that this would bring on the individual. With regard to the Faces sheet the Vietnamese women said that very depressed Vietnamese women would never pick the worst face ('very sad') but would instead choose the slightly milder one ('a bit sad'), again to avoid admitting to having a problem. Vietnamese women also felt that various questions asked by the DIS were culturally inappropriate. Arabic women were concerned with Q15, Q16 and Q18 on the GHQ ('felt you couldn't overcome your difficulties'; 'been finding life a struggle all the time'; and 'been taking things hard'). For the EPDS the Arabic women also considered Q10 to be inappropriate ('the thought of harming myself had occurred to me'); for the Faces sheet they would have preferred only three faces to choose from ('very happy', 'so-so' and 'very unhappy'); and they found several of the DIS questions to be inappropriate.

Of the two other studies, English women interviewed by Poole *et al.*[168] also found that Q10 of the EPDS ('the thought of harming myself has occurred to me') was problematic, and several women in the Cubison and Munro study[169] commented on the overall negative nature of each of the six individual questions included in the shortened, but unvalidated version of the EPDS.

Women commonly expressed concerns about offering truthful answers to questions about depression, the main reason being a fear of disclosure because of shame or of being perceived as an 'incompetent mother'. This issue was raised in four of the seven studies.[167–170] Matthey *et al.*,[170] for example, found that Anglo-Celtic women did not consider any of the questions on the four instruments would prevent women from responding openly. In contrast, Vietnamese women found that certain questions would not elicit true feelings from depressed women because of the shame that disclosure would bring upon them. For Arabic women, certain questions were inappropriate because of the gradations between possible responses. Some English women in the Poole *et al.*[168] study found it difficult to disclose information because of the possible consequences of admitting the truth. As a result of these fears some women reported not being entirely truthful when completing the EPDS or expressed the view that other women might also not be truthful. Shakespeare and colleagues[167] also discovered from their in-depth interviews that many women felt that PND was a stigmatising illness which they would not or did not want to admit to themselves.[167] Some women covered up their feelings for fear of being 'found out' or of losing their baby. Moreover, some women lied deliberately on the questionnaire. Finally, Cubison and Munro[169] found that five of the 15 women that they interviewed admitted to lying when completing the EPDS, and most others commented on the difficulties of being honest.

Interpersonal relationships with health professionals

Three studies[167–169] highlighted the importance of interpersonal relationships between women and health professionals. For example, Poole *et al.*[168] found that the pre-existing relationship with the person administering the questionnaires affected how they were completed. For most women the relationship was with the health visitor and was described as equal and supportive, which helped to increase the likelihood of honesty and disclosure on the part of the mother when completing the questionnaire. The women interviewed by Shakespeare *et al.*[167] suggested that the health visitor should take time and be professional and empathetic about the process of completing the questionnaire. Women reported being able to sense if the health visitor was short of time or uninterested, which gave the impression that completing the questionnaire was just another item to tick off the list of things to do. Women in the Cubison and Munro study[169] suggested that it could be clearer whether the health visitor was an agent of social control and had training in mental

health. Overall, women wanted health professionals to be interested in their emotional well-being in the postnatal period and to be aware of the risk of PND.

Cultural or ethnic differences

Three studies[117,124,170] explored concerns about the cultural or ethnic sensitivity of PND identification strategies. Matthey *et al.*[170] found that Anglo-Celtic women did not consider any of the questions on the four instruments to be culturally inappropriate, whereas for Vietnamese women in particular there was concern about how certain questions would not elicit true feelings for fear of shame. In contrast, when Clarke[124] explored the opinions about postnatal instruments of indigenous First Nations and Métis women of Canada, 90 out of 97 (93%) indicated that they were culturally sensitive and appropriate for how women might feel following the birth of a baby. Furthermore, out of a sample of 34 mothers, 30 (88%) did not feel offended while completing the questionnaires two (6%) indicated that 'may be' they were offended and the remaining two (6%) reported that they did feel offended. When Werrett and Clifford[117] interviewed 23 women to ascertain their views about completing an English and Punjabi version of the EPDS, both were found to be acceptable, but there were mixed responses concerning which version they preferred.

Acceptability of the identification strategy

Three[117,124,167] of the seven studies asked women whether or not they found completing questionnaires acceptable. Shakespeare *et al.*[167] reported that 21 of the 39 women (54%) they interviewed found completing the EPDS to be unacceptable. The three themes that explained the unacceptability of completing the EPDS were: (1) the process of administering the questionnaire; (2) the personal intrusion of this process; and (3) stigma about PND. In contrast, Clarke,[124] in an exploration of First Nations and Métis women's views, found that 86 (97.7%) of 88 women indicated that the questionnaires and interview were appropriate for how women might feel following the birth of a baby. In addition, Werrett and Clifford[117] found that, overall, when women completed English and Punjabi versions of the EPDS, respondents felt comfortable. It was perceived as self-explanatory, understandable and easy to use.

Quantitative synthesis

Gemmill and colleagues,[171] in the only survey of the seven studies, found that of 467 Australian women

who responded to the question using a 5-point Likert scale, 379 women (81%) indicated that completing the EPDS ranged from 'comfortable' to 'very comfortable'. The distributions of responses were no different for those women surveyed by post, by telephone or in a face-to-face interview ($p = 0.49$). Of the 478 women who answered the question, 462 women (97%) thought it was a good idea to assess all new mothers for PND. Women in this survey were also asked what it was like to complete the EPDS. From coding 77 transcripts it was found that the only common category of response to this survey question was that 50 women (65%) thought that completing the EPDS was easy, good or fine.

Integrating qualitative and quantitative synthesis

All seven studies[117,124,167–171] asked women for their views about the EPDS. The majority of the qualitative studies interviewed around 15–40 women, whereas the survey had responses from 479 women. The qualitative studies addressed various themes concerning the method of administration, difficulties in answering questions, interpersonal relationships with health professionals, and cultural or ethnic differences. In contrast, the survey only asked three questions about what it was like to complete the EPDS, how comfortable did women feel in completing the EPDS, and whether women thought it was desirable to complete the EPDS for the identification of PND. The respondents to the survey did not raise any of these themes as being important to them and overall found completing the EPDS to be a comfortable and desirable process. There is evidence from the qualitative studies that women overall found completing the EPDS to be appropriate[124] or comfortable.[117] However, the study by Shakespeare and colleagues,[167] which was the only qualitative study to use in-depth interviews, found that 54% of women thought that health visitors using the EPDS to identify PND was unacceptable.

Health professionals' views of postnatal identification strategies

Table 21 shows that two studies explored the views of health professionals about postnatal identification strategies. Both studies used the EPDS, collected data using semistructured interviews and were conducted in primary care. Health visitors or nurses were recruited using convenience sampling. There was little information describing the sample of health professionals, although the sample size was provided for both

TABLE 21 Health professionals' views of postnatal identification strategies

Study, country, grey literature?	Aim, instruments, version	Design, study setting	Method of sampling, sample	Administration of strategy: timing, setting, mode	Timing of data collection, type of analysis	Conclusions
Poole and Mason, 2008[172] England Yes	To examine health professionals' experiences of using the EPDS to identify PND EPDS English	Semistructured interviews Primary care trust	Convenience 12 practising health visitors, seven nurses	Between 6 and 14 weeks postnatally Home or clinic Unclear	Unclear Interpretative phenomenological analysis	The overall view of using the EPDS was positive and helped to open discussion about PND
Brown and Bacigalupo, 2006[173] England No	To determine health visitors' identification of PND and implications for practice EPDS English	Semistructured interviews Primary care trust	Convenience Six health visitors, experience from 10 months to 10 years	6 weeks postnatally Home or antenatal/birth visits Unclear	Unclear Thematic analyses	The relationship between the client/health visitor is significant, as is experience of training in using the EPDS

EPDS, Edinburgh Postnatal Depression Scale; PND, postnatal depression.

studies. The timing of administration varied from 6 to 14 weeks postnatally and the setting was in women's homes or clinics, but it was not clear how the instruments were administered. Neither study reported when the data on health professionals' views were collected, and interpretative phenomenological and thematic analyses were employed.

Qualitative synthesis
Method of administration of postnatal depression identification strategies

Poole and Mason,[172] in their interviews of health professionals' experiences of identifying PND using the EPDS, found that most of them informed the mothers at early visits that they would be administering the EPDS. It was generally felt that doing this emphasised to mothers that it was routine and helped to normalise the identification of PND and make it acceptable to them. Brown and Bacigalupo,[173] when interviewing six health visitors in their study, also found that health visitors thought that the subject of PND should be raised early on at the antenatal or birth visit.

Poole and Mason[172] found that all but one of the 19 health professionals administered the first routine EPDS at the mothers' homes. Most of the staff voiced strong opinions on the need to complete the EPDS in the home setting. The reason given was that the clinic setting was too busy and mothers could be rushed, whereas at home they were more comfortable. Brown and Bacigalupo[173] also found that home visits were the most appropriate venue to discuss the subject of PND with women, as there was often more time available and it was assumed that women generally felt more comfortable on their own territory if discussing issues concerning emotional health.

In the study by Poole and Mason[172] the health professionals administered the first EPDS at different times. This seemed to be a consequence of both pragmatism and attitude. Timing ranged from 6 weeks to 14 weeks postnatally, although most administration of the EPDS took place around 8 weeks postnatally or soon after. Four health visitors administered the EPDS at 6 weeks postnatally. Staff who administered the EPDS around 12–14 weeks postnatally spoke of having a busy workload. However, staff were satisfied with the later administration date because they felt that at 6 weeks postnatally women were still adjusting to having a baby. All staff undertook a second completion of the EPDS at 8 months postnatally; however, four of the 19 staff thought that this

was too late and that it would be more beneficial at around 4–6 months. Brown and Bacigalupo[173] found that visiting patterns varied from three to six contacts in the first 8 weeks following the birth of the child, and the number of visits depended on the individual mother.

Following completion of the EPDS most health professionals in the Poole and Mason study[172] would go over some or all of the questions with the mother and discuss her answers to ascertain the reasons for particular responses.[172] Sometimes a score was given, but more usually the result was described as high, medium, or low and the implications discussed with the mother as a two-way conversation. It was felt that discussing the results gave the health professionals and mothers the opportunity to discuss issues and let the mothers know that everything was alright. Sharing the results of the EPDS allowed mothers to see their progress using an external scale, but if improvements were not evident then this may have been a negative experience for the mothers.

Difficulties in answering questions

In the Poole and Mason study[172] three of the 19 staff interviewed thought that some of the questions in the EPDS were ambiguous. The last question about self-harm elicited most comments. This related to whether self-harm meant cutting yourself or suicide and the implications of dealing with a positive response. In particular, one member of staff disliked this question saying that it shocked a lot of new mothers. Some staff also experienced difficulties with the question about sleeping, as mothers might not have been able to sleep because of being woken by the baby. In contrast, other staff thought that the EPDS was easy to understand and complete. The health visitors in the study by Brown and Bacigalupo[173] did not express views about difficulties in answering questions when completing the EPDS.

Interpersonal relationships with health professionals

Poole and Mason[172] found that the majority of health professionals' relationships with women improved over a number of meetings, when it became more likely that a mother would disclose additional information. It was suggested that if the relationship with the mother was not going well then discussing the EPDS could improve this. The staff interviewed for this study also found that a woman's presentation was important for informing practice and their relationship with the woman. As an example, non-verbal cues, in conjunction

with clinical experience, might aid a health professional's decision to assess for PND earlier. In contrast, the use of cues might be misleading as a mother can appear to be bubbly and jolly, but then score high on the EPDS for depression. Brown and Bacigalupo[173] also found that the relationship between the health visitor and the mother was significant, with over half the sample referencing the importance of educating mothers about their role. To build a relationship with a woman several contacts are required, beginning antenatally, as this provides an opportunity to compare a women's usual mood with how they present postnatally. Establishing a good relationship with women was important for the early detection of PND. When PND was identified then a number of visits were offered. Health visitor experience was felt to increase the confidence of health visitors in supporting women on an individual basis, but it had to be years of experience of dealing with PND. It also depended on whether it was the right time for a woman to talk about her emotional well-being.

Training issues

In the Poole and Mason study[172] all but two of the health professionals received in-house training in the identification of PND using the EPDS. Five more staff spoke of having assessed women with the EPDS before being trained in its use. The staff did not appear to have a problem using the EPDS before they received training as no negative comments were made in this respect. In addition, some staff felt that they needed more general training to enable them to deal with issues arising around depression and so four undertook a counselling course to give them the skills to deal with the disclosures that their work uncovered. Three other health professionals who had not undertaken additional training reported that sometimes they felt out of their depth in dealing with mothers with depression. However, this opinion was not held by all staff as some who did not do extra training did not express the need for it. Brown and Bacigalupo[173] found in their study of six health visitors that training in the use of the EPDS was minimal. Most health visitors had developed their knowledge of PND through self-directed study involving searching the internet and reading appropriate community nursing journals. To better prepare themselves in identifying and supporting women with PND all health visitors felt that they would benefit from further and more consistent training, which should include all members of the primary health-care team.

Acceptability of identification strategy

Poole and Mason[172] found that all staff were positive about using the EPDS, and the EPDS was seen as a tool that opened up discussion around PND. The feelings of the health visitors towards the EPDS in the Brown and Bacigalupo study[173] were more mixed. One health visitor felt that the questionnaire allowed a mother to report her feelings more objectively, which would produce richer information than had the same points been raised through general discussion. Another respondent felt that the EPDS was useful because it was a different kind of prompt in face-to-face discussion than if the subject of PND had been raised in general terms. However, another respondent felt that the EPDS was 'overused and open to manipulation'.

Women's views of prenatal identification strategies

Three studies explored the views of women about prenatal identification strategies as shown in *Table 22*. Two studies assessed the EPDS and a third study assessed a standardised questionnaire to predict women with PND. Interviews were used as the method of data collection in all three studies. One study was conducted in an antenatal clinic, another in general practice and for one study the setting was unclear. Women were recruited from convenience samples with the number ranging from 15 to 202. English-speaking women were included in two studies (one of which also included Arabic- and Vietnamese-speaking women) and French women only in another study. The instruments were administered at varying times antenatally, including the first, second and third trimester. In two studies the instruments were administered in women's homes and in one study an antenatal clinic. The timing of data collection varied across three trimesters, and both narrative synthesis and thematic analysis were used to synthesise evidence.

Qualitative synthesis

The results from the three studies are consistent with the findings on women's views about postnatal identification strategies. It was again found that self-completion of a questionnaire at home helped to avoid feelings of intrusion and made it easier for women to respond to questions honestly.[176] Difficulties were also found in answering questions from the EPDS about self-harm[174] as well as the question about sleeping – it was thought that all pregnant women have difficulty sleeping and that this might not be the result of being unhappy.[175]

TABLE 22 Women's views about prenatal identification strategies

Study, country, grey literature?	Aim, instruments, version	Design, study setting	Method of sampling, sample	Administration of strategy: timing, setting, mode	Timing of data collection, type of analysis	Conclusions
Matthey et al., 2005[174] Australia No	To assess acceptability of routine use of instruments antenatally EPDS English	Semistructured telephone interviews Antenatal clinic	Convenience 104 English, 50 Vietnamese and 48 Arabic women	First antenatal visit Antenatal clinic Interview by clinic midwife	First antenatal visit and 5–8 weeks postnatally Narrative synthesis	The EPDS was found to be acceptable by the majority of English- and non-English-speaking women
Clark, 2000[175] England No	To assess the suitability of use of the EPDS by health visitors EPDS English	Semistructured interviews General practice	Convenience 15 women and three health visitors	Third trimester Home (two clients at antenatal clinic) Self-completed questionnaire	Third trimester Thematic analysis	EPDS facilitates discussion between clients and health visitors about emotional health
Jacquemain and Golse, 1998[176] France No	To develop a questionnaire that will help health professionals predict women with PND Questionnaire French	Semistructured interviews Unclear	Convenience 30 women; mean age 31 years; 29 married and one single	24 weeks (second trimester) Home Self-completed questionnaire	8 days after completing questionnaire Narrative synthesis	No-one was made to feel uncomfortable by the questions and completing at home avoided feelings of intrusion

EPDS, Edinburgh Postnatal Depression Scale; PND, postnatal depression.

67

With regard to interpersonal relationships, Matthey and colleagues[174] found that the majority of English women felt comfortable with the midwife asking them questions about their psychosocial health. They also highlighted that most English, Arabic and Vietnamese women found the EPDS acceptable.

Health professionals' views of prenatal identification strategies

Two studies presented in *Table 23* explored the views of health professionals towards prenatal identification strategies, one qualitative[175] and the other quantitative.[177] The EPDS was used in both studies and one study also included a pregnancy questionnaire, which is a 17-item PI for measuring vulnerability to PND. Interviews or a survey using a 5-point Likert scale were used as the methods of data collection. Both studies were conducted in primary care and included a convenience sample of health visitors. The standardised questionnaires were administered in the third trimester as self-completed questionnaires. In one of the studies data about health professionals' views were collected in the third trimester and in the other study the timing of data collection was unclear. The qualitative study used thematic analyses and the survey reported health visitors' responses as percentages.

Integrating qualitative and quantitative synthesis

Both studies were limited in their discussion of the topics identified as important from the research literature. Clark's[175] interviews of health visitors addressed issues about the method of administration of the EPDS and interpersonal relationships, and found that the EPDS should be introduced earlier in the pregnancy and was best administered at home and that, although the EPDS facilitated 'opening up', it was easier to use with women with whom health visitors already had a relationship. The survey conducted by Wood[177] addressed the overall acceptability of the identification strategies and discovered that for both the EPDS and the pregnancy questionnaire the majority of health visitors agreed that the questionnaires allowed sensitive issues to be raised in a structured format and that they would use these types of questionnaire in practice.

Women's views of perinatal identification strategies

Table 24 presents three studies[178–180] that explored the views of women about perinatal identification

strategies. All three studies used the EPDS and data were collected using interviews, focus groups and a survey in either an antenatal clinic or maternity service setting. Women were recruited into the studies using convenience, purposive or random sampling; 10 and 12 women were included in the qualitative studies and 860 in the survey. The EPDS was administered at varying times antenatally and at around 6–8 weeks postnatally. The setting and mode of administration of the EPDS in two studies were unclear, but for one study the setting was in a hospital and for another study the EPDS was completed as a questionnaire or interview. Women's views were collected postnatally in two studies, but the timing of administration was unclear for the third study. Narrative synthesis was used in the two qualitative studies and the percentages of women responding to the instrument in the survey study.

Integrating qualitative and quantitative synthesis

Both qualitative studies[178,180] addressed issues of difficulty in answering individual items of the EPDS (anxious or worried, scared or panicky, self-harm) and in particular voiced concerns about completing the questionnaire honestly. In addition, Alder[180] briefly addressed issues about interpersonal relationships and cultural or ethnic differences. It was found that the EPDS was easy to complete and one route into dialogue with the health visitor but that what mattered was the warmth and intimacy of the relationship. The small group of ethnic minority women had problems with the absence of any culturally contextualised understanding or awareness of PND. The survey of Buist and colleagues[179] did not address any of the themes that were identified from the research literature but, similar to the qualitative studies, concluded that the majority of women found the EPDS easy to complete and experienced no discomfort.

Health professionals' views of perinatal identification strategies

There were two studies[179,180] that explored health professionals' views of perinatal identification strategies as shown in *Table 25*. Both studies used the EPDS in a maternity services setting and data were collected using either focus groups and interviews or a survey. A convenience sample of health professionals was selected for each study with 16 in the qualitative study and over 200 GPs, nurses or midwives in the survey. In both studies the EPDS was administered antenatally and at around 6–8 weeks postnatally. In one study the EPDS was administered in a hospital, but the setting was unclear for the other study. It was also

TABLE 23 Health professionals' views about prenatal identification strategies

Study, country, grey literature?	Aim, instruments, version	Design, study setting	Method of sampling, sample	Administration of strategy: timing, setting, mode	Timing of data collection, type of analysis	Conclusions
Clark, 2000[175] England No	To assess the suitability of health visitors' use of the EPDS EPDS English	Semistructured interviews General practice	Convenience Three health visitors	Third trimester Home (two clients at antenatal clinic) Self-completed questionnaire	Third trimester Thematic analyses	EPDS facilitates discussion between clients and health visitors about emotional health
Wood, 2002[177] Scotland Yes	To assess health visitors' perceptions of the utility of these questionnaires EPDS, pregnancy questionnaire English	Survey Primary care trust	Convenience 35 health visitors	Between 32 and 36 weeks of pregnancy Unclear Self-completed questionnaire	Unclear Report responses as percentages	Over two-thirds (69%) of health visitors did not feel uncomfortable with the content of either questionnaire

EPDS, Edinburgh Postnatal Depression Scale; PND, postnatal depression.

TABLE 24 Women's views of perinatal identification strategies

Study, country, grey literature?	Aim, instruments, version	Design, study setting	Method of sampling, sample	Administration of strategy: timing, setting, mode	Timing of data collection, type of analysis	Conclusions
Edge, 2005[178] England No	To assess women's views about whether the EPDS was a valid tool for detecting PND EPDS English	In-depth interviews Antenatal clinics	Purposive 12 Black Caribbean women	Last trimester and 6 weeks postnatally Unclear Self-completed questionnaire or interview	Between 6 and 12 months postnatally Narrative synthesis based on prespecified themes	Women thought that the EPDS was a valid and acceptable indicator of depression
Buist et al., 2006[179] Australia No	To assess acceptability of routine identification of perinatal depression EPDS English	Survey Maternity services	Random sample of women 860 women (average age 31 years)	Antenatally (16, 28 or 36 weeks); postnatally (6–8 weeks) Hospital Unclear	Between 8 and 12 weeks postnatally Report responses as percentages	Over 90% of women found the EPDS easy to complete
Alder, 2007[180] Scotland Yes	To explore the views of service users and health professionals on use of the EPDS EPDS English	Focus groups and interviews Maternity services	Convenience 10 women (five from ethnic minority groups)	Antenatally (28 weeks); postnatally (6 weeks) Unclear Unclear	Unclear Narrative synthesis	Majority of women found the EPDS easy to complete

EPDS, Edinburgh Postnatal Depression Scale; PND, postnatal depression.

TABLE 25 Health professionals' views of perinatal identification strategies

Study, country, grey literature?	Aim, instruments, version	Design, study setting	Method of sampling, sample	Administration of strategy: timing, setting, mode	Timing of data collection, type of analysis	Conclusions
Buist et al., 2006[179] Australia No	To assess the acceptability of routine identification of perinatal depression EPDS English	Survey Maternity services	Convenience GPs (n = 229, 20 years' experience); nurses (n = 267, 19 years' experience); midwives (n = 305, 16 years' experience	Antenatally (16, 28 or 36 weeks); postnatally (6–8 weeks) Hospital Unclear	Between 8 and 12 weeks postnatally Report responses as percentages	The majority of health professionals (71–83%) found the EPDS easy or fairly easy to complete
Alder, 2007[180] Scotland Yes	To explore the views of service users and health professionals on use of the EPDS EPDS English	Focus groups and interviews Maternity services	Convenience 16 health professionals (health visitors, community midwives, voluntary sector staff)	Antenatally (28 weeks); postnatally (6 weeks) Unclear Unclear	Unclear Narrative synthesis	Health visitors found the EPDS useful to help raise the subject of PND

EPDS, Edinburgh Postnatal Depression Scale; PND, postnatal depression.

unclear for both studies whether the EPDS was used as a self-completed questionnaire or not. In one study health professionals' views were collected postnatally and in the other study the timing of data collection was unclear. The qualitative study used narrative synthesis and the survey presented responses as percentages.

Integrating qualitative and quantitative synthesis

In the exploration of health professionals' use of the EPDS by Alder[180] the findings were consistent with those of other studies, such as introducing the EPDS antenatally at 28 weeks to alert mothers to the possibility that they might experience mood disturbance and being aware that the questionnaire might not be answered honestly and so should be used to open up discussion. The survey[179] did not address any themes and instead reported that most health professionals [84 (71%) GPs, 190 (83%) maternal child health nurses and 147 (76%) midwives] found the EPDS easy to complete.

Discussion

Women liked to be informed in advance that they would be asked to complete a questionnaire to identify PND.[167,168] If they were not prepared they might have felt anxious and reluctant to answer questions honestly. Health professionals also felt that informing mothers about the completion of questionnaires at earlier visits helped to normalise the process of identification of PND and make it acceptable to them.[172,173] They suggested that this subject should be raised early on at an antenatal visit.[173,175] Having raised the subject of PND identification there was variation in when women and health professionals thought was the appropriate time to undertake this. Women said that it should be at around 8 weeks postnatally, as a period of adjustment to being a mother was required.[168] Some health professionals agreed with the need for mothers to adapt to having a baby and thought that PND identification should be undertaken at around 6–8 weeks postnatally,[172,173] although others felt that it should be introduced during pregnancy.[175,180] There was consensus, however, between women[167,168,176] and health professionals[172,173,175] that questionnaires should be administered at the woman's home. Administering questionnaires in clinics was found to be too distracting and uncomfortable for women, whereas at home there was more privacy and time to discuss issues concerning emotional health, which made it easier to respond to questions more honestly.

In general, when administering the instrument women preferred to talk rather than complete a standardised questionnaire, and they were critical of the lack of dialogue that could result from a pen and paper assessment.[167–169] Similarly, in general, women appreciated being given feedback of the results of completing the questionnaire as they then felt listened to and found it helpful to talk freely with the health visitor.[167,168] Health professionals also found that following the completion of a questionnaire it was useful to discuss the results and engage in a two-way conversation.[172]

Several studies made criticisms about specific questions used in the instruments to identify PND. For the EPDS the last question about the thought of self-harm was identified as inappropriate for reasons including the extent of shame that this would bring on an individual in some cultures.[170,174] English women also reported feeling vulnerable and not wanting to admit to self-harm[168] and black Caribbean women found the item difficult to answer because of the uncertainty about whether it referred to accidental or deliberate self-harm.[178] Similarly health professionals expressed difficulties with this question because of the ambiguity about 'cutting' or attempting suicide.[172] English women and health professionals also found difficulties with the question about sleeping as mothers might not have been able to sleep because of being woken by the baby.[172,174] More generally, women expressed concerns about offering truthful answers to questions about depression for fear of disclosure and losing their baby or the shame of being seen as an incompetent mother.[167–170] Moreover, women in four studies[167,169,178,180] reported that they deliberately lied on the questionnaire for fear of answering questions honestly. In contrast, some health professionals thought that the EPDS was easy to understand and complete[173] or did not identify specific questions as sensitive or difficult to ask.[180]

As mothers found it helpful to talk about their feelings and difficult to answer questions honestly, the interpersonal relationship with the health professional was important. Indeed, two studies found that the pre-existing relationship with the person administering the questionnaire affected how it was completed.[168,175] Most mothers wanted the health professional to be supportive and caring and show an interest,[167,168,174] which helped increase the likelihood of honesty and disclosure.[167] Furthermore, the relationship between mothers and health professionals improved over a number of meetings, when it became more likely that a

mother would disclose additional information,[172] and so several contacts antenatally helped to establish a good relationship for the early detection of PND.[173] Although an instrument such as the EPDS was seen as a route into dialogue with a health professional,[172,180] what mattered was the warmth and intimacy of the relationship.[180] Women and health professionals also found it important to be clear about the role of the health professional so that mothers would not perceive them as agents of social control.[169,173] Adequate training in identifying PND would therefore appear necessary to facilitate the relationship between mother and health professional. Most health professionals received in-house training in PND identification using the EPDS, but had no problems using the EPDS before training. In contrast, some staff sought training in counselling to acquire the skills for dealing with the disclosures that their work uncovered.[172] Health visitors also had very little training in PND, which was mostly self-directed study using the internet or reading journals, and thought that they would benefit further from more consistent training.[173] Therefore the perceived need for training ranges from not requiring any training to specialised training in counselling. Training health visitors in the appropriate use of the EPDS and non-directive counselling skills has also been shown to reduce EPDS scores.[181]

Some concerns raised about the acceptability of the EPDS were culturally specific.[178] Compared with Anglo-Celtic women, who did not consider questions culturally inappropriate, Vietnamese women felt that questionnaires used to identify PND would not elicit true feelings for fear of shame.[170,174] Black and minority ethnic women were also reported as having problems with the absence of any culturally contextualised understanding or awareness of PND,[180] and black Caribbean women found some items in the EPDS difficult to answer, although other items were good indicators of depression.[178] In contrast, indigenous First Nations and Métis women of Canada found questions in instruments to be culturally sensitive and appropriate and were mostly not offended.[124] English and Punjabi women also found completing the EPDS to be acceptable.[117]

Both qualitative and quantitative methods were used to address the overall question of whether or not postnatal identification strategies were acceptable to women and health professionals. Most qualitative studies[117,174,180] showed that English-speaking women thought that the EPDS was acceptable, although one study[167] that

used in-depth interviews found that 21 of 39 English-speaking women felt that the EPDS was unacceptable. This study has limitations, though, in that the interviews were undertaken several months after the completion of the EPDS and thus the study may have selected women who preferred to talk rather than complete forms. Nearly all First Nations and Métis women found questionnaires to be appropriate, and black Caribbean women overall found the EPDS to be acceptable.[178] From the perspective of the health professionals, qualitative studies showed that staff were positive about using the EPDS,[172] although while it could be a useful prompt in face-to-face discussion[173,180] it could be overused and open to manipulation.[173] Two surveys including several hundred Australian women found that the majority indicated that PND identification using the EPDS was 'comfortable' to 'very comfortable'[171] or that the EPDS was fairly easy to complete.[179] A survey[177] of health visitors also found that the majority would use the 17-item pregnancy questionnaire and the EPDS in practice and most did not feel uncomfortable with the content of the questionnaire. Another survey[179] of 801 health professionals (GPs, nurses and midwives) found that the majority thought that the EPDS was easy or fairly easy to complete by their patients and found no discomfort in explaining the EPDS.

For our study a systematic approach was used to underpin the synthesis of the evidence base concerning the acceptability of postnatal identification strategies to women and health professionals. We identified several studies by undertaking a comprehensive literature search, independently selected eligible studies and extracted data, and synthesised data using the textual narrative approach. We also collected data from qualitative studies, which examined in detail the views of a small sample of women or health professionals on the acceptability of identification strategies, and from surveys, which quantified the acceptability of these strategies for samples of several hundred. There are limitations, however, to the validity of the findings and the generalisability of the review.

First, it was difficult to integrate the evidence from qualitative and quantitative research because the surveys only asked broad questions about, for example, how comfortable women found the process of PND identification. It was not possible to assess how different themes discussed in qualitative research studies were included in the surveys and how subsequently this affected the estimates of acceptability of identification strategies to

women and health professionals. Therefore, from surveys alone it was not possible to understand what made an identification strategy acceptable or not. Furthermore, the surveys could be open to selection bias in that only one[179] of the four surveys selected a random sample. The surveys also had poor response rates further indicating the possibility of bias in those women or health professionals who chose to respond and possibly limiting the validity and generalisability of the findings. The response rates were 52% for a survey of Australian women,[171] 29% for a survey of Scottish health visitors,[177] and 59%, 20%, 50% and 29%, respectively, for a survey of Australian women and GPs, nurses and midwives.[179] No surveys were conducted on women other than those in Australia.

Second, although qualitative research methods are a more useful design to help understand the acceptability of identification strategies from women's and health professionals' perspectives, it is important to critically appraise the study designs. Most studies recruited a convenience sample of participants and collected data using semistructured interviews. Subsequently, the sampling strategy did not promote the generalisability of the individuals included in the sample. Moreover, the two studies[167,178] that used in-depth interviews to explore this subject were the most critical about the acceptability of postnatal identification strategies using the EPDS. Most studies collected data around the same time as the process of PND identification, but a criticism of both studies which used in-depth interviews is that the data were collected several months after the process of PND identification. This suggests that there might be recall bias in the response of the participants or that a particular type of participant wanted to take part in the study when collecting data several months later. However, it might be that women's views about the acceptability of PND identification soon after the administration of, for example, the EPDS might reflect the relief that labour is over and that it takes more negative aspects of this process longer to integrate. This has been found with women after childbirth, whose assessments of their experiences can change from positive to negative over time.[182]

Third, most studies that explored women's views about PND identification strategies did report their EPDS scores. This helps with understanding the prevalence and severity of PND in the sample. The findings, however, were often not presented and subsequently discussed in the context of whether the women had PND or not, which

might influence how acceptable they found the identification strategy. For example, Buist *et al.*[179] found from their survey of women who completed a questionnaire about the acceptability of the EPDS that 87% of women who had an EPDS score < 13 experienced no discomfort in completing the EPDS compared with 64% of women with an EPDS score > 13, which was statistically significant ($p < 0.0001$). Furthermore, the EPDS and other self-completed questionnaires are not incontrovertible as women will be misclassified as having PND or not (i.e. false positives and false negatives). None of the studies explored women's views about the acceptability of an identification strategy when this misclassification occurred.

Fourth, there are a number of identification strategies for identifying PND, such as the use of standardised or generic questionnaires postnatally or prenatally, the identification of known risk factors through prenatal identification of PND, or the use of training packages to enhance the awareness of health professionals about the clinical signs of PND. Our review only identified studies of pre- or postnatal identification strategies using standardised or generic questionnaires, and every study but one used the EPDS as the instrument of choice. Although the preponderance for the use of the EPDS reflects clinical practice, this has highlighted an evidence gap in the assessment of the acceptability of alternative identification strategies. In particular, it is possible that strategies such as prenatal identification using known risk factors or training packages targeted at health professionals could avoid the need for a paper and pen assessment and issues surrounding women answering these questions openly and honestly.

Reflection on current policy and practice within the UK

As outlined in Chapter 1, recent NICE guidance recommends the use of case-finding questions to help identify women with PND [(1) 'During the past month, have you often been bothered by feeling down, depressed or hopeless?', (2) 'During the past month, have you often been bothered by little interest or pleasure in doing things?' and (3) 'Is this something you feel you need or want help with?']. Whilst these questions appear to offer a relatively quick and convenient way for healthcare professionals to identify post-natal depression, the acceptability of these questions remains unexplored. Some of the findings from this review would indicate that the case finding

questions may potentially be acceptable to women and health professionals. The finding that women wanted to talk rather than complete a paper and pencil questionnaire, would potentially support the use of the case findings questions over using the EPDS. Furthermore, the case finding questions may overcome some of the difficulties surrounding specific items on the EPDS (e.g. question 10) and may be less culturally sensitive. The findings that women and health professionals should be forewarned of the use of formal methods to identify PND, that they should be administered in the woman's' home and the importance of the interpersonal relationship between the woman and the health professional may still be important factors in the acceptability of the case finding questions. Further research into the acceptability of these difference identification strategies is desirable.

Chapter 7

Clinical effectiveness of methods to identify postnatal depression in improving maternal and infant outcomes: systematic review 3

In phase 3 we reviewed whether the routine use of case identification strategies for PND or the integration of case identification strategies with enhancements of care resulted in improvements in maternal and infant outcomes.

Methods of review

Searching

Studies were primarily identified from the database searches outlined in Chapter 3. A comprehensive systematic review was undertaken as part of the antenatal and postnatal mental health guidance issued by NICE in October 2007.[30] A number of systematic reviews regarding prevention and treatment of antenatal and postnatal depression have also been published under the auspices of the Cochrane Collaboration.[183–188] Reference lists of all systematic reviews on PND identification were scrutinised to identify any additional studies for inclusion in this review.

Inclusion criteria

Two reviewers screened titles and abstracts generated by the searches to identify potentially eligible studies. Any disagreements were resolved by consensus or deferred to a third party if necessary. Full papers of potentially eligible studies were obtained and assessed for inclusion

independently by two reviewers. Articles were eligible for inclusion if they fulfilled the criteria outlined below.

Study design

The review included RCTs and controlled trials. Trials had to include an identification strategy component that was incorporated in some way into clinical decision-making. Hence, trials that included an identification strategy to gain a baseline measure of depression and did not use the results from the identification strategy for any other purpose were excluded from the review. A hierarchy of evidence was established that incorporated different models of assessment of the effectiveness of PND identification strategies on maternal and infant outcomes. For the purposes of this review we considered the highest level of evidence to come from trials comparing methods to identify PND with no formal methods or delayed methods to identify PND (*Figure 25*).

There is some suggestive evidence from a review by the US Preventative Services Task Force[189] that using identification strategies for depression can become effective when they are accompanied by organisational enhancements of care, involving clinician education, support from case managers and a collaborative care approach between specialists and primary care physicians.[190] Hence, this systematic review examined the impact of using
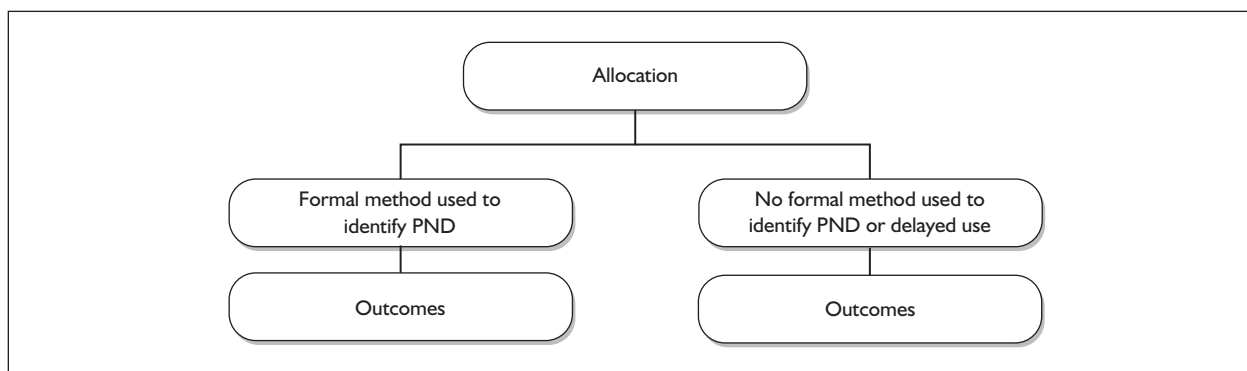
FIGURE 25 Level I evidence.

identification strategies by themselves and also identification strategies that had been incorporated into some from of educational or organisational enhancement of care.

The first model of level II evidence (top of *Figure 26*) compares using an identification strategy plus some form of enhancement of care with usual care. Hence, participants in the control group would not be formally assessed using the PND identification strategy. The second model (bottom of *Figure 26*) involves the whole sample of participants receiving the identification strategy, but randomising them so that the intervention group receive feedback of their results from a health-care professional and the control group do not receive feedback of their results. Some form of enhancement of care could also be incorporated into this model.

Finally, it was anticipated that there would be relatively few, if any, trials that would fit the above two models; hence, broader criteria were developed and level III evidence incorporated trials that used the PND identification strategy as an eligibility criterion in a trial (*Figure 27*). For example, women scoring 12 or above on the EPDS would be included in the trial and randomised to the intervention group or usual care group. Women scoring 11 or below on the EPDS would not be eligible for inclusion in the trial. Trials comparing more than one intervention were also included as long as there was a usual care group to which women could be allocated.

Trials that could be categorised into one of the three levels of evidence described above were eligible for inclusion in phase 3 of the review.
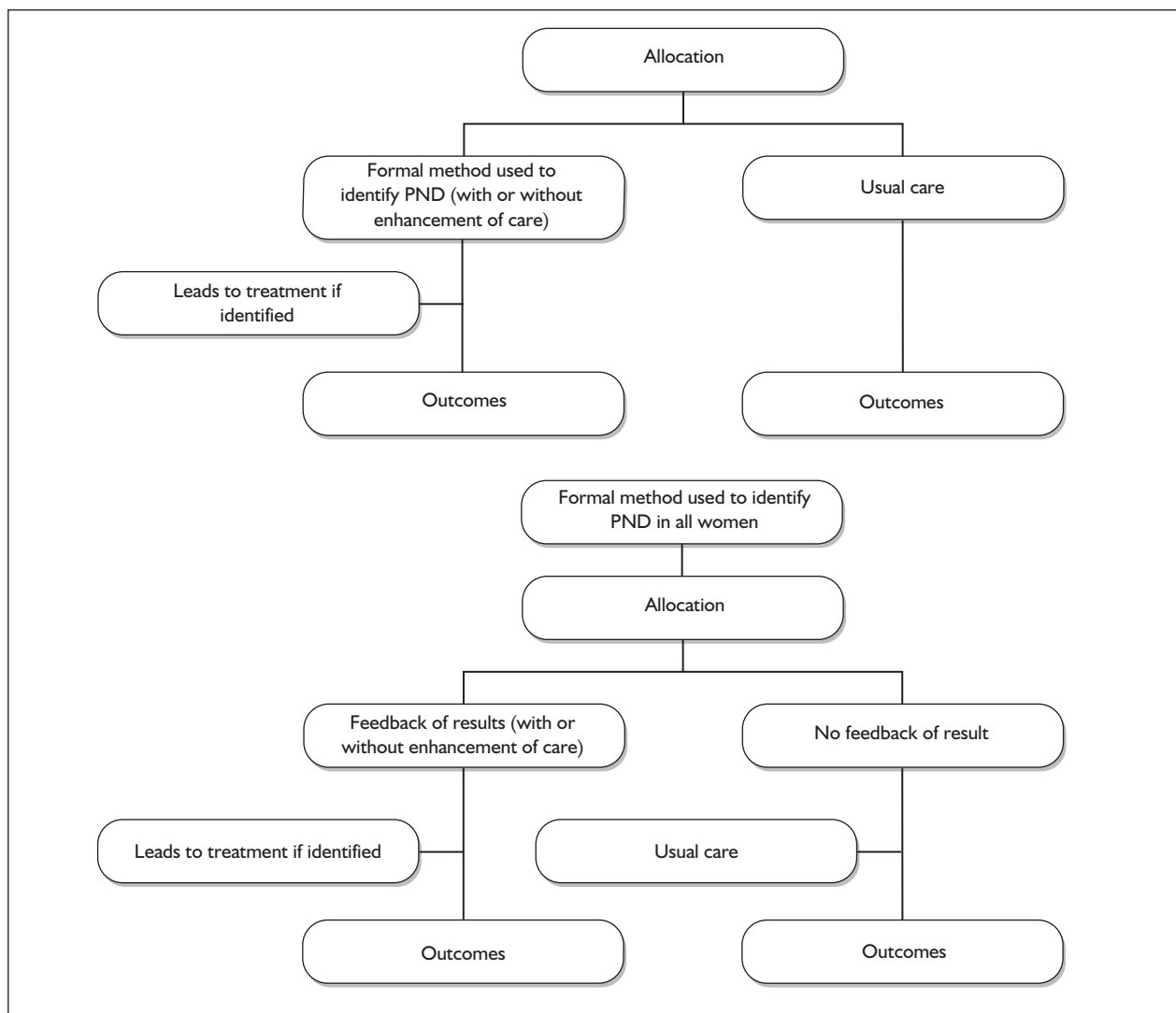
FIGURE 26 *Level II evidence.*
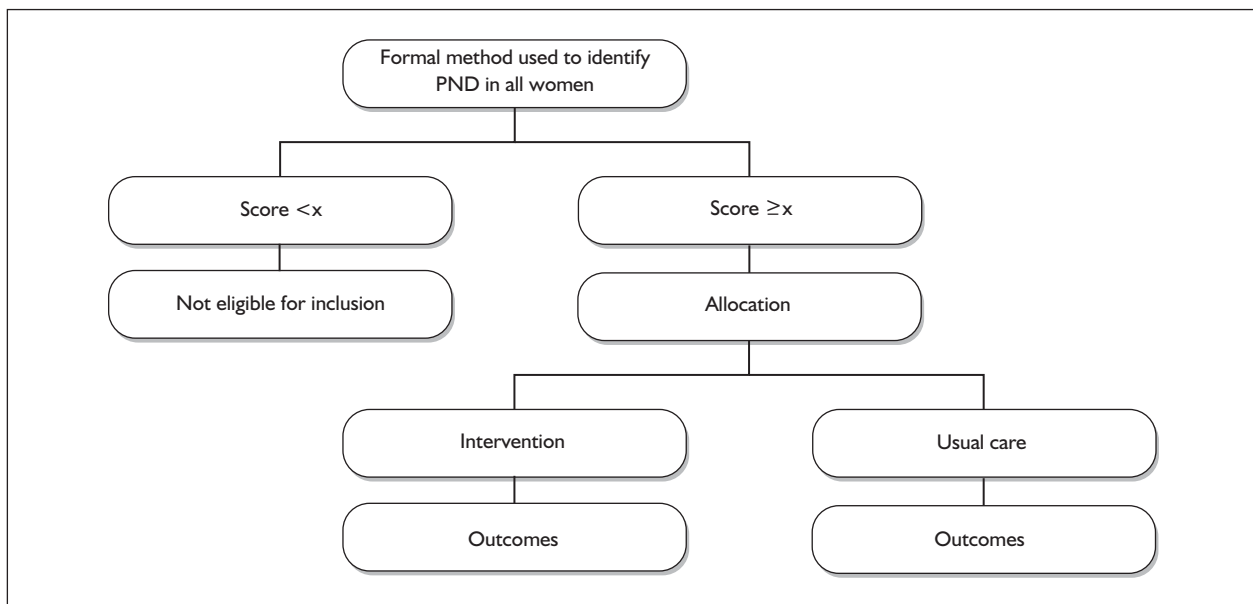
FIGURE 27 *Level III evidence.*

Interventions

Trials comparing the implementation of any PND identification strategy, with or without enhancement of postnatal care, in any setting were included. Any type of intervention was included as long as it was compared with usual primary care for expectant and postnatal mothers. Head-to-head trials comparing interventions without the option of participants being randomised to a usual care group (e.g. a trial comparing antidepressants with placebo) were excluded from the review. Trials assessing interventions aimed at health-care professionals and/or at women with PND were included in the review. Studies providing level III evidence were divided into prevention and treatment studies, and were categorised into four types: educational, psychosocial, psychological and more complex interventions that included educational and/or psychosocial and/or psychological components. Prevention studies were those that recruited women and delivered the intervention under study while women were pregnant, with the aim of preventing the onset of PND. Treatment studies recruited women and delivered the intervention under study during the postnatal period (any time immediately following birth until 1 year postnatally).

Outcomes

Data on maternal and infant outcomes were included, however defined. Short-term (< 6 months), medium-term (6–12 months) and longer-term (> 12 months) outcomes were all considered. Trials reporting no maternal or infant outcomes were excluded from the review.

Data extraction

Data were extracted by one reviewer using a predefined data extraction form. The following data were extracted: author, year of publication, study design, setting, patient population, details of the intervention and usual care, sample size and results. When there were multiple publications for the same study, data were extracted primarily from the most recent and complete publication. In cases in which the duplicate publications reported additional relevant data, these data were also extracted.

Data synthesis

For level I and level II evidence, for each dichotomous outcome, the numbers of patients experiencing the outcome were extracted for each group. The odds ratio (OR) and 95% CI were calculated for each study outcome. When there was more than one study for a comparison, the ORs were pooled using a fixed-effect model [the Mantel–Haenszel (M–H) method] and the corresponding 95% CIs were calculated. Statistical heterogeneity was assessed using the I^2 statistic, as outlined in Chapter 5. It was intended that continuous data would be analysed by calculating the difference in means and corresponding 95% CIs for each study. To perform a meta-analysis of continuous data we needed to extract mean values of the outcome, the standard deviation and the number of participants included in the analysis of the outcome data. In the three studies reporting continuous outcome data it was impossible to obtain standard deviations for each group from

the data summaries presented in each publication. Hence, formal pooling was not undertaken for the continuous outcome data.

For level III evidence no statistical pooling was performed because of clinical heterogeneity between the studies. The study results are presented in a narrative synthesis, grouped by level of evidence and by the type of study (prevention or treatment of PND) and identification strategy used. Graphical presentations of dichotomous and continuous outcomes are displayed in forest plots to allow inspection of the data; however, overall pooled analyses were not undertaken and hence pooled estimates are not included on the forest plots. Odds ratios and 95% CIs were calculated for each study providing dichotomous outcomes and differences in means and corresponding 95% CIs were calculated for each study providing continuous outcomes. All analyses were conducted using STATA version 9.[100]

Results

Overall, 30 studies (from 32 publications) were included in phase 3 of the review. Only five of these studies provided level I or level II evidence, with the remaining 25 studies providing level III evidence. Among the studies categorised as providing level III evidence there were four main types of intervention under study: educational, psychosocial, psychological and more complex interventions that included educational and/or psychosocial and/or psychological components. Fuller details of the included studies are provided in Appendix 5.

Level I and level II evidence

Five studies[191–195] were identified that provided level I or level II evidence. One of the studies compared the rate of identification of PND using the EPDS with the rate of spontaneous detection from routine clinical evaluations by physicians and mid-level health-care providers in a residency training programme.[191] Women were recruited from a single hospital in the USA and allocated to groups based on the date of delivery; those women who delivered on even days were assigned to the EPDS group. Women assigned to the EPDS group were posted the EPDS to complete at 6 weeks postnatally. Data for the control group were gathered by review of the postnatal record. Only 37% ($n = 79$) of women in the intervention group returned the EPDS and only 54% ($n = 96$) of the control group women

attended the 6-week postnatal visit. The rate of detection of PND was 35.4% ($n = 28$) in the EPDS group and 6.3% ($n = 6$) in the routine clinical evaluation group ($p < 0.0001$). A cut point of 10 was used to indicate a high risk for PND.

The remaining studies focused on detecting women experiencing PND and subsequently reducing the number experiencing PND or the severity of PND (indicated by a reduction in identification scores), rather than detecting PND alone as in the previous study. All of the studies used the EPDS (alone or in combination with other strategies) as an identification strategy, although the EPDS was administered at various time points: 25 weeks' gestation;[192] 4 weeks postnatally;[193] 1 day before discharge, 1 week postnatally and 6 weeks postnatally;[194] or unclear.[195] The EPDS was used as an outcome measure across all of the studies, either as a continuous variable comparing mean EPDS scores between the intervention and control groups or as a dichotomous variable comparing the number of women scoring above or below a cut point on the EPDS. Outcomes were assessed at 36 weeks' gestation;[192] 6 weeks postnatally[194] or 4 months postnatally.[193,195] Additional outcomes to be considered were number of visits to midwife and GP ($n = 2$), physical and mental components of the SF-36 ($n = 1$), number of referrals ($n = 1$), women's views about care ($n = 1$), overall satisfaction with care ($n = 1$) and the use of a support contact number ($n = 1$). The results of each individual study for the threshold EPDS scores and the mean EPDS scores are displayed in *Figures 28* and *29* respectively. For the studies providing dichotomous outcomes the I^2 value was 0%.

The aim of the study by Wickberg *et al.*[192] was to compare the antenatal management of depression when midwives were aware of women's EPDS scores with the management of depression when midwives were not aware of women's EPDS scores. Women were recruited from antenatal care centres in six sectors of primary health-care districts in Sweden. Midwives were randomised to the intervention and control groups, rather than the individual women themselves. Additional training about depression was given to midwives in the intervention group. In total, 669 women took part in the study, 318 in the intervention group and 351 in the control group. Women completed the EPDS twice, first at gestational week 25 and second at gestational week 36. Those women scoring > 11 on the EPDS in the intervention group were phoned and asked for permission to inform their midwife about the EPDS score. Analysis showed that in gestational
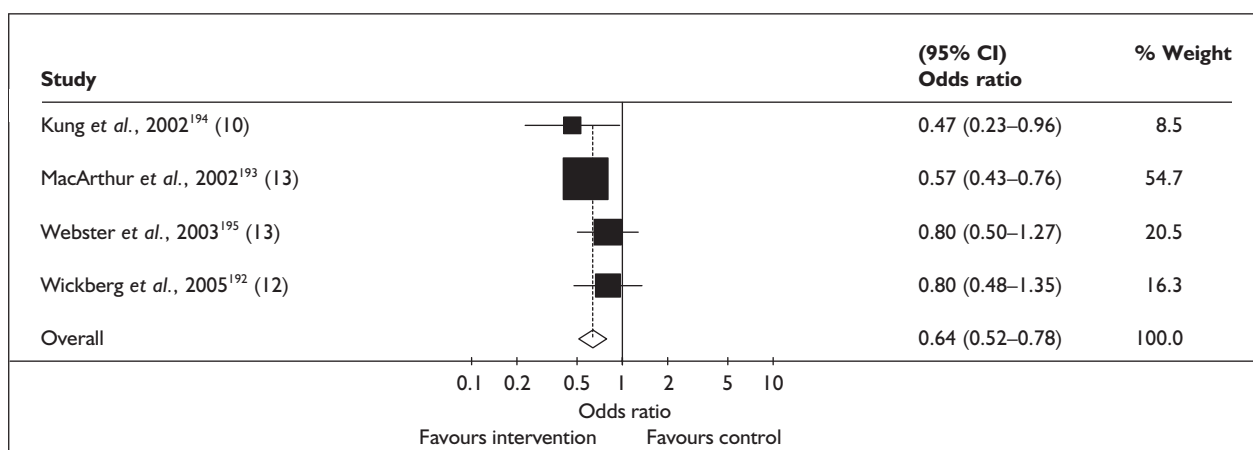
FIGURE 28 *Forest plot showing odds ratios for EPDS threshold scores.*
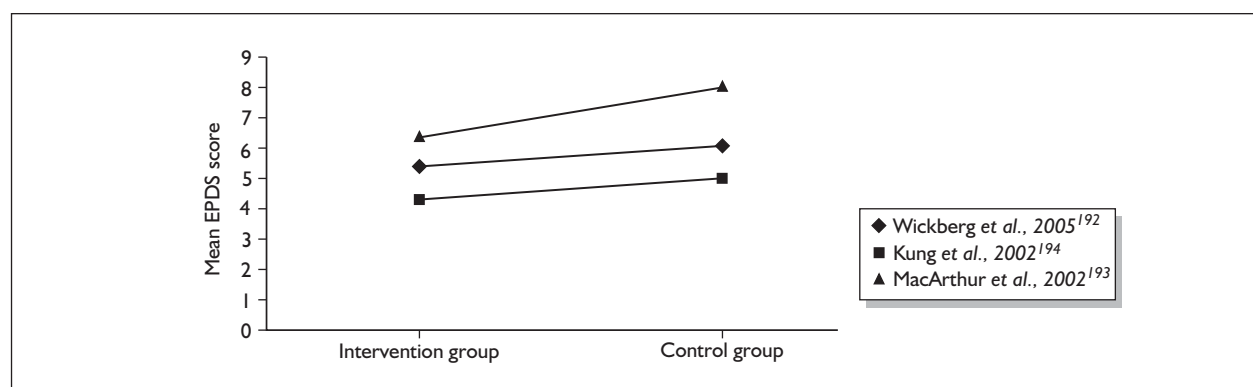
FIGURE 29 *Summary of mean EPDS scores in the intervention and control groups.*

week 25, 48 (15%) of the 318 women in the intervention group and 45 (13%) of the 351 women in the control group scored above 11 on the EPDS, whereas in gestational week 36, 26 (10%) of the 273 women in the intervention group and 40 (12%) of the 345 women in the control group scored above the threshold. This difference was statistically significant ($p < 0.0001$). Women in the intervention group also displayed a reduction in the mean EPDS score from week 25 (6.41, range 0–25) to week 36 (5.39, range 0–19) ($p < 0.001$); however, the mean EPDS score in the control group remained almost the same (6.07, range 0–21, versus 6.11, range 0–22). None of the analyses accounted for the clustering by midwife.

A second study by Webster *et al.*[195] evaluated the effectiveness of a prenatal intervention in reducing the incidence of PND. Women attending their first prenatal appointment at a single hospital in Australia were screened for risk factors of PND [Maternity Social Support Scale (MSSS) < 25; past personal or family history of mental illness or PND]. Those women identified with elevated risk

were randomised to the intervention ($n = 299$) or control group ($n = 301$). The intervention consisted of a booklet about PND with contact numbers, use of the EPDS, a discussion with the woman about her risk of developing PND and a letter to the woman's GP and local child health nurse alerting them to the woman's risk of PND (i.e. feedback of the risk). Women in the control group received standard care, which included case management and referral to a hospital social worker or psychiatrist if required. Women were asked to complete a second EPDS at 16 weeks postnatally. Of the 509 women sent the second EPDS, 73% ($n = 371$) responded. The number of women with an EPDS score of 12 or above was 24% ($n = 46$) in the intervention group and 28% ($n = 50$) in the control group (OR 0.80, 95% CI 0.50 to 1.28).

Another study[194] investigated whether the provision of an early intervention by midwives decreased the incidence of depression at 6 weeks postnatally. The sample of women was recruited from a single teaching hospital in Hong Kong. Women in the intervention group were assessed with the EPDS

three times during the postnatal period (1 day before discharge, 1 week postnatally and 6 weeks postnatally). The intervention included an in-depth assessment by midwives (women were asked to describe any sleep disturbance, poor appetite or self-blame or if they had ideas of hurting the baby), with a subsequent phone follow-up service by midwives and/or volunteers to the women with high EPDS scores (cut point 10). Training on basic counselling and interviewing skills and emotional problems of postnatal women was provided to midwives, social workers and volunteers before the study commenced. Women in the control group were not assessed with the EPDS until 6 weeks postnatally. In the intervention group 12 (5.9%) women had EPDS scores above 9 at 6 weeks postnatally compared with 24 (11.8%) women in the control group ($p = 0.03$). The mean EPDS score of the intervention group was 4.32 whereas that of the control group was 4.97; an exact p-value was not reported in the article – it was reported only that the p-value was not statistically significant.

The final study[193] explored the effects of redesigning community postnatal care on women's health 4 months postnatally. A total of 36 general practices from the West Midlands were cluster randomised using minimisation to the intervention or control group. General practices were randomised rather than women to prevent contamination. In the intervention group, care was led by midwives and included extended care to 28 days, use of a symptoms checklist and the EPDS at day 28, referral to a GP as necessary, and a 10- to 12-week discharge visit. The control group received routine care that included seven midwifery home visits until 10–14 days postnatally (although this could be extended to 28 days) and care by health visitors thereafter. GPs completed routine home visits and a final check-up at 6–8 weeks postnatally. Additional training in postnatal care and health and trial design was provided to all midwives. In total, 17 practices (1087 women) were randomised to the intervention group and 19 practices (977 women) were randomised to the control group. The mental health score was significantly better in the intervention group than in the control group (OR 3.03, 95% CI 1.53 to 4.52; OR_{adj} 4.31, 95% CI 2.50 to 6.12), as were the mean EPDS score (OR –1.92, 95% CI –2.55 to –1.29; OR_{adj} –2.68, 95% C –3.46 to –1.89) and the proportion of women with an EPDS score of 13 or above (OR 0.57, 95% CI 0.43 to 0.76; OR_{adj} 0.47, 95% CI 0.31 to 0.76). The physical health score did not differ between study groups (OR –0.79, 95% C –1.91 to 0.34; OR_{adj} –0.80, 95% C –2.32 to 0.72). Overall satisfaction

with care from the community midwives did not differ statistically between groups (OR 1.09, 95% CI 0.72 to 1.63).

Level III evidence

In total, 25 studies were identified that provided level III evidence, nine of which were prevention studies (*Table 26*) and 16 of which were treatment studies (*Table 27*). There were a number of clinical and methodological differences between the studies. Most importantly for the focus of this review there were differences between the studies in terms of the identification strategy used during recruitment, the cut point chosen and the timing of administration of the identification strategy. The studies also differed with respect to the intervention used, the outcome assessed, the timing of the outcome measure and any subsequent follow-up measures. Because of the heterogeneity between studies included in the review it was not possible to combine the data to produce an overall pooled estimate; however, individual study estimates are still presented in forest plots to allow visual inspection of the data.

For the prevention studies a variety of identification strategies were used: BDI ($n = 3$), Crown–Crisp Experimental Index (CCEI) ($n = 2$), EPDS ($n = 1$), GHQ ($n = 1$), Modified Antenatal Screening Questionnaire (MASQ) ($n = 1$), PI ($n = 1$) and risk factors ($n = 1$). The identification strategies were administered at various time points during pregnancy, ranging from within 18 weeks' gestation to 32 weeks' gestation. Forest plots for dichotomous and continuous outcomes are displayed in *Figures 30* and *31* respectively. In the forest plots the studies are grouped by the identification strategy used during recruitment and ordered within these groups by the cut point used, which is reported in parentheses after the publication details. All but one of the prevention studies reporting dichotomous outcomes reported the number of women with depression as defined by a diagnostic interview;[196,198–200,204,205] the final study[203] used the EPDS at a cut point of 13 to classify women as being depressed or not. Interestingly, this last study is the study located on the opposite side of the forest plot to all of the other studies. We can clearly see that there is wide variability across the results in these studies.

For the continuous outcomes four[197–200] of the studies reported mean EPDS scores, with the remaining three[201,204,205] studies reporting mean BDI scores. Two[197,198] of the studies reporting mean

TABLE 26 *Prevention studies*

Study	Intervention	Identification strategy (cut point), timing	Threshold depression	Continuous depression
Brugha et al., 2000[196]	Preparing for Parenthood – structured antenatal risk factor-reducing intervention designed to increase social support and problem-solving skills	Pregnancy and You plus any one of six depression items on GHQ-D 12–20 weeks' gestation	EPDS, GHQ, ICD-10 (3 months)	
Buist et al., 1999[197]	Intervention group classes	Eight or more risk factors 12–24 weeks' gestation		EPDS, BDI (6 and 24 weeks)
Elliot et al., 1988[198]	Psychoeducation and social support in a programme with continuity of care	Leverton questionnaire: poor marital relationship or previous psychological problems or score of more than 10 on the CCEI anxiety items Pregnancy (within 18 weeks' gestation)	Bedford College (3 months)	EPDS, CCEI-depression, CCEI-anxiety (3 months)
Elliot et al., 2000[199]	Social support, education, continuity of care	Scored 2 on any one of the vulnerability questions in Leverton questionnaire or scored 1 on more than one question or scored ≥ 10 on the CCEI anxiety subscale During pregnancy	Bedford College (3 months)	EPDS (3 months)
Gorman, 2001[200]	Interpersonal therapy	Definite personal history of major depressive disorder with or without treatment; BDI (13); first-degree relative with a history of psychiatric treatment; DAS < 100; experiencing two or more life events predicted to have moderate or severe negative impact for the woman in the postnatal period Third trimester	DSM-III-R (1 and 6 months)	BDI, SCL-90-R, EPDS (1 and 6 months)
McKee et al., 2006;[201] Zayas et al., 2004[202]	Cognitive behavioural therapy + child development psychoeducational modules + social support building sessions	BDI (14) Within 32 weeks' gestation		BDI-II (2 and 12 weeks)
Stamp et al., 1995[203]	Support groups	MASQ (2) 24 weeks' gestation	EPDS (6, 12 and 24 weeks)	
Zlotnick et al., 2001[204]	Interpersonal therapy	At least one risk factor: previous history of depression; poor social support; BDI > 10; recent stressful event 20–32 weeks' gestation	DSM-IV (3 months)	BDI (3 months)
Zlotnick et al., 2006[205]	Interpersonal therapy	Predictive Index (27) 23–32 weeks' gestation	Longitudinal interval follow-up interview (3 months)	BDI (3 months)

BDI, Beck Depression Inventory; CCEI, Crown–Crisp Experimental Index; DAS, Dyadic Adjustment Scale; DSM, *Diagnostic and Statistical Manual of Mental Disorders*; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; ICD, *International Classification of Diseases, Injuries and Causes of Death*; MASQ, Modified Antenatal Screening Questionnaire; SCL-90-R, Symptom Checklist 90-Revised.

TABLE 27 Treatment studies

Study	Intervention	Identification strategy (cut point), timing	Threshold depression	Continuous depression
Armstrong and Edwards, 2003[206]	Exercise and social support	EPDS (12) 6 weeks to 12 months postnatally	EPDS (6 and 12 weeks)	EPDS, DASS, GHQ (6 and 12 weeks)
Chabrol et al., 2002[215]	Cognitive behavioural therapy	EPDS (9) 2–5 days postnatally	HAMD, BDI, EPDS (10–12 weeks)	HAMD, BDI, EPDS (10–12 weeks)
Chen et al., 2000[214]	Support groups	BDI (10) 3 weeks postnatally	BDI (10–14 weeks)	BDI (10–14 weeks)
Cooper et al., 2003;[216] Murray et al., 2003[217]	Cognitive behavioural therapy, psychodynamic therapy, non-directive counselling	EPDS (12) Early postnatally	DSM-III-R (4.5 and 9 months)	EPDS (4.5 and 9 months)
Dennis, 2003[207]	Peer support	EPDS (10) 8 weeks postnatally	EPDS (4 and 8 weeks)	EPDS (4 and 8 weeks)
Holden et al., 1989[218]	Counselling	EPDS (12) 6 weeks postnatally	RDC (3 months)	EPDS, psychiatric interview (3 months)
Heh and Fu, 2003[208]	Informational support – posted PND booklet	EPDS (10) 4 weeks postnatally	EPDS (3 months)	EPDS (3 months)
Honey et al., 2002[209]	Psychoeducational group intervention: cognitive behavioural therapy + education + relaxation	EPDS (13) Postnatally	EPDS (8 weeks and 6 months)	EPDS (8 weeks and 6 months)
Horowitz et al., 2001[221]	Coached behavioural intervention to promote maternal–infant responsiveness	EPDS (10) 2–4 weeks postnatally		BDI-II (10–14 weeks and 14–18 weeks)
Ingadottir and Thome, 2006[210]	Web-based course for health professionals	EPDS (12) 9 weeks postnatally	EPDS (15 and 24 weeks)	EPDS (15 and 24 weeks)
Meager and Milgrom, 1996[220]	Education, social support and cognitive behavioural therapy	EPDS (13), BDI (16) 6 months postnatally		BDI, EPDS, POMS (10 weeks)
Milgrom et al., 2005[222]	Cognitive behavioural therapy or counselling	DSM-IV diagnosis 6–18 weeks postnatally		BDI (3 months)

Study	Intervention	Identification strategy (cut point), timing	Threshold depression	Continuous depression
Misri et al., 2004[211]	Cognitive behavioural therapy	HAMD (18), HAMA (20), EPDS (12) Postnatally	HAMD, HAMA, EPDS (3 months)	HAMD, HAMA, EPDS (3 months)
O'Hara et al., 2000[213]	Interpersonal therapy	Inventory to diagnose depression, DSM-IV, HAMD (12) Postnatally	HRSD, BDI (4, 8 and 12 weeks)	HAMD, BDI (4, 8 and 12 weeks)
Prendergast and Austin, 2001[212]	Cognitive behavioural therapy	EPDS (13) Mean = 15 weeks postnatally	EPDS (6 weeks and 6 months)	EPDS (6 weeks and 6 months), MADRS (6 weeks)
Wickberg and Hwang, 1996[219]	Counselling	EPDS (12) 2 months postnatally	DSM-III-R (6 weeks)	MADRS (6 weeks)

BDI, Beck Depression Inventory; DASS, Depression Anxiety Stress Scale; DSM, *Diagnostic and Statistical Manual of Mental Disorders*; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; HAMA, Hamilton Rating Scale for Anxiety; ; HAMD, Hamilton Rating Scale for Depression; HRSD, Hamilton Rating Scale for Depression; MADRS, Montgomery–Asberg Depression Rating Scale; PND, postnatal depression; POMS, Profile of Mood States; RDC, Research Diagnostic Criteria.
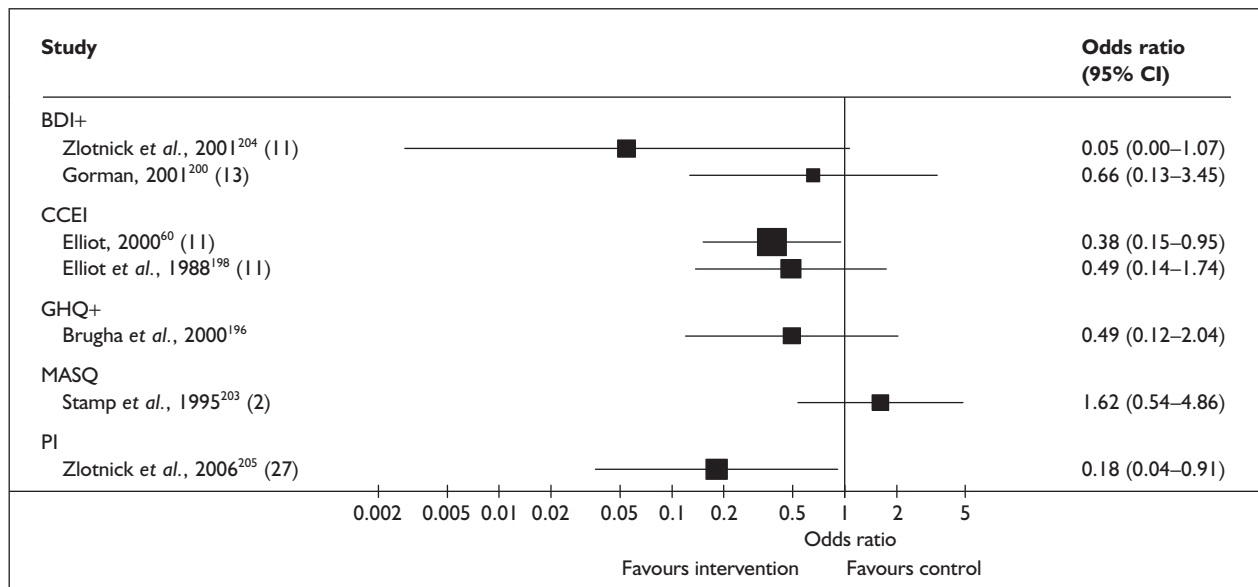
FIGURE 30 *Forest plot of prevention studies reporting various dichotomous outcomes.*
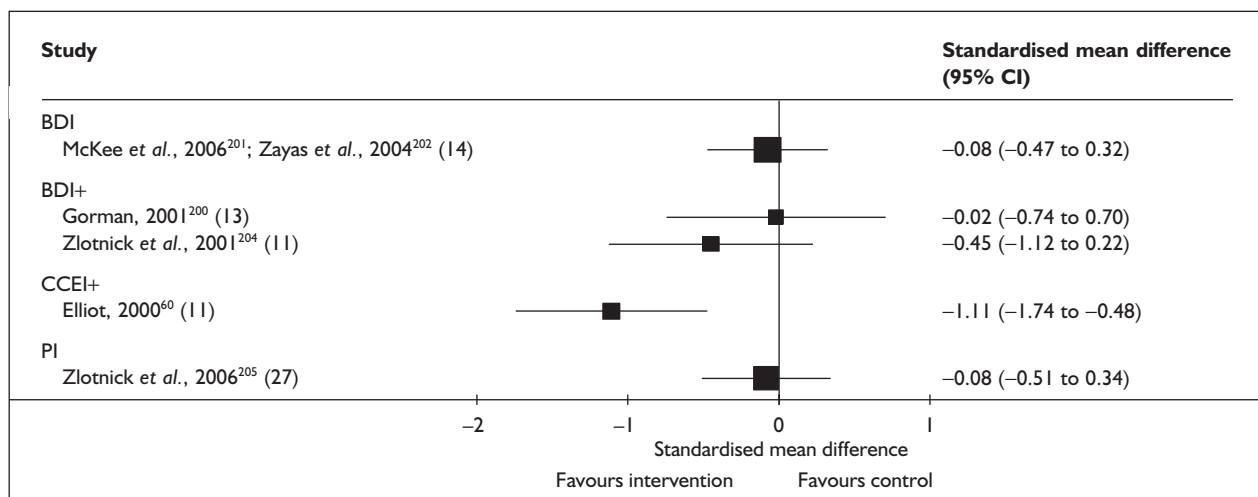
FIGURE 31 *Forest plot of prevention studies reporting various continuous outcomes.*

EPDS scores did not report measures of dispersion and hence could not be displayed in the forest plot. We can clearly see that there is wide variability across the results in these studies.

In total, 13 of the treatment studies used the EPDS (alone or in combination with other strategies) during the recruitment stage of the study. The studies used different cut points on the EPDS to distinguish between those women who were at risk of having PND and those who were not. Cut points of 9, 10, 12 and 13 were used. Using different cut points leads to different types of women being included in the studies. The remaining treatment studies used generic depression identification strategies (BDI or HRSD) or diagnostic interviews. Identification strategies were administered at various time points during the first year postnatally,

ranging from 2–5 days postnatally to 12 months postnatally; however, the majority of women in the studies were administered the identification strategies within 6 months postnatally. Forest plots for dichotomous and continuous outcomes are displayed in *Figures 32* and *33* respectively. Seven[206–212] of the treatment studies reported dichotomous outcomes as the number of women with depression as defined by the number of women scoring about a cut point on the EPDS, three[213–215] studies used the BDI and the final three[216–219] studies used diagnostic interviews. For the continuous outcomes, 11 of the studies reported mean EPDS scores,[206–212,215–218,220] with the remaining four studies reporting mean BDI scores.[213,214,221,222] Three of the studies reporting mean EPDS scores[210,218,220] and two reporting mean BDI scores[214,222] did not report measures of

Study	(95% CI) Odds ratio
BDI	
Chen *et al.*, 2000[214] (10)	0.33 (0.12–0.96)
EPDS	
Chabrol *et al.*, 2002[215] (9)	0.01 (0.00–0.13)
Dennis, 2003[29] (10)	0.16 (0.04–0.72)
Heh and Fu, 2003[208] (10)	0.31 (0.11–0.82)
Wickberg and Hwang, 1996[219] (12)	0.75 (0.15–3.87)
Holden *et al.*, 1989[218] (12)	0.27 (0.08–0.86)
Cooper *et al.*, 2003[216]; Murray *et al.*, 2003[217] (12)	0.81 (0.39–1.67)
Ingadottir and Thome, 2006[210] (12)	0.21 (0.02–2.47)
Armstrong and Edwards, 2003[206] (12)	0.02 (0.00–0.50)
Honey *et al.*, 2002[209] (13)	0.31 (0.08–1.16)
Prendergast and Austin, 2001[212] (13)	0.36 (0.03–3.85)
EPDS+	
Misri *et al.*, 2004[211] (12)	0.57 (0.13–2.48)
HRSD+	
O'Hara *et al.*, 2000[213] (12)	0.36 (0.16–0.83)

FIGURE 32 *Forest plot of treatment studies reporting various dichotomous outcomes.*

Study	Standardised mean difference (95% CI)
EPDS	
Chabrol *et al.*, 2002[215] (9)	−2.37 (−3.12 to −1.61)
Horowitz *et al.*, 2001[221] (10)	0.85 (0.48–1.23)
Dennis, 2003[29] (10)	−0.86 (−1.49 to −0.22)
Heh and Fu, 2003[208] (10)	−0.35 (−0.82 to 0.13)
Armstrong and Edwards, 2003[210] (12)	−1.71 (−2.75 to −0.67)
Cooper *et al.*, 2003[216]; Murray *et al.*, 2003[217] (12)	0.02 (−0.31 to 0.35)
Prendergast and Austin, 2001[212] (13)	−0.37 (−1.06 to 0.32)
Honey *et al.*, 2002[209] (13)	−0.51 (−1.15 to 0.13)
EPDS+	
Misri *et al.*, 2004[211] (12)	−0.20 (−0.94 to 0.54)
HRSD+	
O'Hara *et al.*, 2000[213] (12)	−1.10 (−1.52 to −0.67)

FIGURE 33 *Forest plot of treatment studies reporting various continuous outcomes.*

dispersion and hence could not be displayed on the forest plot. We can clearly see that there is wide variability across the results for both dichotomous and continuous outcomes

Discussion

Five studies were identified that compared the use of an identification strategy with or without enhancement of care or feedback of scores with not using an identification strategy or usual care (i.e. providing level I or II evidence). As identified in previous chapters, the EPDS was the most frequently used identification strategy. All of the studies indicated beneficial effects of using the EPDS in reducing EPDS scores, although some of the individual studies did not show statistically significant differences. Studies reporting dichotomous outcomes were combined and the pooled estimate gave an OR of 0.64 (95% CI 0.52 to 0.78). Thus, the odds of

scoring above the threshold for depression in a population in which a formal method was used to identify PND (with an intervention for those identified) were 0.64 times the odds of scoring above the threshold for depression in a population in which there was no formal method to identify PND (with an intervention for those identified). Problems occurred when attempts were made to combine studies reporting continuous outcomes as measures of dispersion were often not reported. As three out of the five studies also included some form of enhancement of care it is impossible to disentangle the effect of the identification strategy component alone. For example, it could be that the enhancement of care is the important factor and use of the identification strategy has little impact.

A total of 25 studies were identified that reported using identification strategies at the recruitment stage to identify women at risk of PND with the aim of examining the impact of various interventions compared with usual care (i.e. level III evidence). Despite a large number of studies being identified there were a number of clinical and methodological differences between the studies, which did not permit statistical pooling of results. Furthermore, it was hard to distinguish between the benefit of using identification strategies and the effects of the intervention under study. Thus, it was difficult to draw conclusions about the impact of using identification strategies on maternal and infant outcomes. Further research would be informative in this area.

There were a number of methodological weaknesses of the studies included in this review. We included controlled trials as well as randomised trials and, for example, in some of the studies odd or even expected dates of delivery were used to randomise participants to treatment groups. Obviously, such methods of randomisation are not truly random and thus, in practice, this often results in a selection bias being introduced. In addition, in some of the studies described as RCTs it was often unclear how the randomisation sequence was generated and what methods were used to conceal the sequence. Hence, it was difficult to judge whether the methods used to randomise participants to treatment groups were subject to bias. In some studies randomisation was undertaken on a cluster rather than on an individual basis. When data are ordered in clusters, individuals may not be independent of one another and subsequently this needs to be taken into account in the analysis. In two out of the three studies using cluster randomisation the analyses did not account for the clustering. When clustering is not accounted for in the analysis it results in p-values and standard errors (SEs) that are too low and confidence intervals that are too narrow. Another frequently occurring problem was attrition, with most studies reporting that some women were lost to follow-up. Non-compliance or non-attendance was also reported across studies. It was often unclear whether analyses were undertaken using an intention to treat approach or not. An intention to treat approach should be used as the primary analysis when non-compliance occurs as this ensures that the initial randomisation sequence is preserved.

Reflection on current policy and practice within the UK

As outlined in Chapter 1, current NICE guidance recommends the use of two questions to identify possible depression and a third question if the women answers 'yes' to either of the initial questions [(1) 'During the past month, have you often been bothered by feeling down, depressed or hopeless?', (2) 'During the past month, have you often been bothered by little interest or pleasure in doing things?' and (3) 'Is this something you feel you need or want help with?']. No studies were identified in this review that assessed the clinical effectiveness of these questions in improving maternal and infant outcomes in a postnatal population.

Chapter 8

Cost-effectiveness of methods to identify postnatal depression: systematic review 4

In this phase we reviewed whether the routine use of case identification strategies for PND or the integration of case identification strategies with enhancements of care were cost-effective.

Methods

Inclusion criteria

Two reviewers screened titles and abstracts to identify potentially eligible studies. Any disagreements were resolved by consensus or deferred to a third party if necessary. Full papers of potentially eligible studies were obtained and assessed for inclusion independently by two reviewers.

Articles were eligible for inclusion if they were full economic evaluations (cost–benefit analyses, cost-effectiveness analyses and cost–utility analyses) of PND identification strategies. We followed explicit guidelines laid down by the CRD in the preparation of the NHS EED.[42] The quality and relevance of any available economic data were judged from the perspective of the UK NHS according to criteria laid down by Drummond.[223]

Results

The additional searches for full economic evaluations yielded 3667 articles. On the basis of the title and abstract two studies from the original searches and two studies from the economic searches appeared to be potentially eligible for inclusion. On closer inspection none of the studies were a full economic evaluation of a PND identification strategy. The reasons for exclusion from the review are outlined in *Table 28*.

Discussion

Despite an extensive systematic search of the literature none of the studies identified presented a full economic evaluation of a PND identification strategy. Future research needs to address this gap in the literature. We are aware of one large, but yet unpublished, RCT with an economic evaluation (PoNDER trial). The PoNDER trial aimed to assess the costs and effectiveness and broad impact of two health visitor psychological interventions for PND.

TABLE 28 *Summary of reasons for exclusion*

Study	Comparison	Reason for exclusion
Morrell *et al*., 2000[224]	Community postnatal support workers vs routine primary care	EPDS used as an outcome not as an identification strategy
Boath *et al*., 2003[225]	Psychiatric day hospital vs routine primary care	EPDS used as an outcome not as an identification strategy
Appleby *et al*., 2003[226]	Training health visitors in cognitive behavioural counselling	Not a full economic analysis; only the costs of health visitor time pre- and post-training were considered
Petrou *et al*., 2006[227]	Preventative intervention vs routine primary care	Not a full economic analysis of a PND identification strategy; cost-effectiveness analysis of the intervention under study

EPDS, Edinburgh Postnatal Depression Scale; PND, postnatal depression.

Reflection on current policy and practice within the UK

As outlined in Chapter 1, recent NICE guidance recommends the use of case-finding questions to help identify women with PND [(1) 'During the past month, have you often been bothered by feeling down, depressed or hopeless?', (2) 'During the past month, have you often been bothered by little interest or pleasure in doing things?' and (3) 'Is this something you feel you need or want help with?']. No studies assessing the cost-effectiveness of these questions (or any other method to identify PND) were identified.

Chapter 9

Decision model of methods to identify postnatal depression

In the absence of existing cost-effectiveness studies of PND identification strategies we developed a new decision-analytic model to evaluate the cost-effectiveness of a range of alternative identification strategies. The model provided a framework for the synthesis of diagnostic accuracy data reported in Chapter 5 with a range of other relevant parameters required to establish the cost-effectiveness of using formal identification strategies for women with PND. The model also provided a vehicle for identifying potential future research priorities, reported in Chapter 10.

Key concepts in cost-effectiveness analyses of identification strategies

To inform the development of a decision-analytic model it is important to establish the key features of cost-effectiveness analyses for the purpose of informing resource allocation in the NHS:

1. The specification of the decision problem should ideally include the comparison of all identification strategies that could feasibly be used in the NHS. It is recognised, however, that in practice these options may be constrained by the availability of evidence and the structural complexity of any model.
2. The analysis should make a clear link between the diagnostic accuracy of a given identification strategy, the impact on subsequent treatment decisions and the ultimate effect on health outcomes and costs. Hence, the costs and outcomes of each of the four diagnosis groups – true positive, false negative, true negative and false positive – need to be assessed.
3. The ultimate health effects of the alternative identification strategies should be expressed in terms of a generic measure of health such as a quality-adjusted life-year (QALY). This is because it is necessary to assess the value of improved outcomes from more accurate identification strategies in units that can be compared with those of programmes and

interventions in other specialties and disease areas that are competing for finite health-care resources.
4. The evidence used to estimate cost-effectiveness should be relevant to patients and clinical practice in the UK health service.
5. The uncertainty in the evidence base needs to be reflected in the model. To simultaneously assess the implications of uncertainty in all elements of evidence, probabilistic analysis should be used to establish the decision uncertainty associated with each diagnostic strategy being compared.[228,229] This informs decision-makers about the probability of each strategy being the most cost-effective conditional on the value that the decision-maker places on a unit of health gain. Such methods can also be used to provide an opportunity to apply value of information (VOI) methods to inform priority setting in research.[43,230,231]

Methods

The objective of the model was to estimate, based on best available data, the costs and health outcomes for a range of feasible identification strategies. The analysis was conducted from an NHS/personal social services (PSS) perspective, with costs expressed in 2006/7 prices and health outcomes expressed in terms of QALYs. The time horizon of the analysis was 1 year; as such, no discounting of costs and health outcomes was applied.

The model was made up of two parts including: (1) an identification model reflecting the diagnostic performance and administration costs of the alternative identification strategies; and (2) a treatment model that evaluated the subsequent costs and outcomes (expressed in QALYs) of each of the four diagnosis groups – true positive, false negative, true negative and false positive.

The model was probabilistic in that input parameters were entered into the model as

probability distributions to reflect second order uncertainty – that is, uncertainty surrounding the mean estimates.[229] Monte Carlo simulation was used to propagate uncertainty in input parameters through the model in such a way that the results of the analysis could also be presented with their uncertainty. The probabilistic analysis also provided a formal approach to quantifying the consequences associated with the uncertainty surrounding the model results and could be used to identify priorities for future research. Scenario analysis was used to explore the impact of alternative assumptions in the model.

The following sections outline the decision problem and the structure of the model and also provide an overview of the key assumptions and data sources used to populate the model in more detail. A prerequisite of the treatment model was ensuring that subsequent treatments for PND (i.e. for women with a true positive diagnosis) were, in themselves, cost-effective. Given the importance of this aspect and as this issue was, in part at least, separable from the broader question of the cost-effectiveness of identification strategies, the cost-effectiveness of alternative treatment strategies for the management of women with PND was considered first. This analysis was used to identify the optimal treatment (according to cost-effectiveness considerations) that was assigned to women with PND as part of the broader modelling work.

Establishing the cost-effectiveness of treatment for women with postnatal depression
Overview

For any PND identification strategy to be considered cost-effective, a cost-effective treatment strategy must exist for those women with PND correctly diagnosed as depressed. The recent NICE guidance on antenatal and postnatal mental health identified structured psychological therapy and listening home visits as the most suitable first-line treatments for depression in the postnatal period.[30] A deterministic economic model conducted as part of the NICE guidance found listening home visits to be both more effective and more costly than structured psychological therapy. The incremental cost-effectiveness ratio (ICER) of listening home visits versus structured psychological therapy was reported to be £9435 per QALY. However, neither treatment strategy was compared with usual care (i.e. the management of PND without a formal policy of listening home visits or psychological

therapy); hence, it was not clear from the NICE guidance whether either treatment strategy was cost-effective compared with usual care. To clarify this the NICE treatment model was reconstructed with a usual care arm; the opportunity was also taken to revise some of the parameter values used and to reflect uncertainty in these values using probabilistic methods.

Model structure and key structural assumptions
The reconstructed treatment model took the form of a decision tree and a schematic is given in *Figure 34*. The schematic was adapted from the one presented in the NICE guidance and included a 'usual care' alternative. For a detailed description of the model, reference is made to the NICE guidance;[30] the following provides a very brief summary of the model structure and key structural assumptions.

At the model start point a woman with PND was provided with one of the two treatments (with corresponding 'additional care', as described in the NICE guidance) or with usual care. If a treatment was provided the woman could continue or discontinue treatment, whereas it was assumed that the woman could not discontinue usual care. The woman then either responded to the treatment (or improved under usual care) or did not respond. If the woman responded or improved then it was assumed that she underwent a linear improvement in health-related quality of life (HRQoL), evaluated in terms of QALYs, from the model start point to week 8. If the woman then relapsed within the time horizon of the model it was assumed that she underwent a linear deterioration in HRQoL from week 8 to 1 year (the model end point), otherwise she remained non-depressed until the end point. If the woman did not respond to treatment or improve under usual care it was assumed that she remained depressed until the model end point.

Parameter inputs for the treatment model
The parameter values incorporated into the reconstructed NICE treatment model are reported in Appendix 6. A brief summary of the approaches used to inform the relevant parameters are reported in the following sections.

Risk of discontinuing treatment
The NICE estimate for the absolute risk of discontinuation for usual care was calculated by summing over the control arms of all psychological treatment studies given in the relevant guideline meta-analysis (01.04). A revised value for the
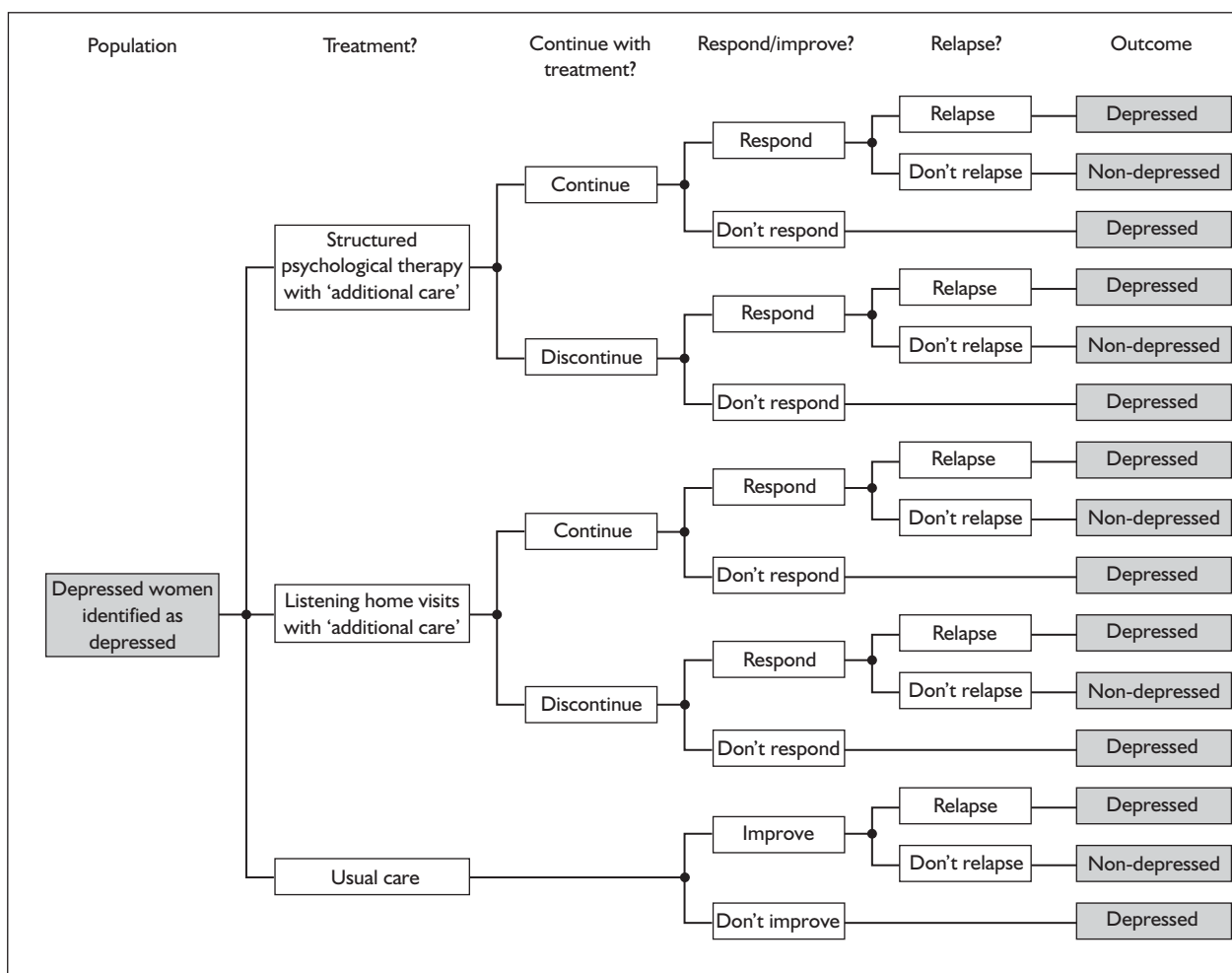
FIGURE 34 *Schematic of the treatment model (adapted from NICE guideline[30]).*

absolute risk of discontinuation for usual care was calculated by carrying out a meta-analysis over the same control arms. As in the NICE guidance,[30] the absolute risk of discontinuation for each treatment was calculated by multiplying the absolute risk of discontinuation for usual care by the relative risk of discontinuation for the respective treatment and then subtracting the absolute risk of discontinuation for usual care. The justification for this was given in the NICE guidance (p. 179), and the calculation remained unchanged here.

Risk of no response/improvement
The NICE estimate for the absolute risk of no improvement for usual care was calculated by summing over the control arms of all psychological treatment studies given in the relevant guideline meta-analysis. A revised estimate for the absolute risk of no improvement for usual care was calculated by undertaking a random-effects meta-analysis over the same control arms. As in the NICE guidance,[30] the absolute risk of no response

from each treatment was calculated by multiplying the absolute risk of no improvement from usual care by the relative risk of no response from the respective treatment (see NICE guidance, p. 179).

Risk of relapse
The NICE estimate for the risk of relapse was assumed to be common for women who responded to treatment and those who improved under usual care and was calculated by summing over the treatment arms of the studies included in the relevant meta-analysis (08.04.06.03) given in Appendix 19b of the NICE guideline on depression.[89] The estimate of the risk of relapse for women who improved under usual care was revised by carrying out a random-effects meta-analysis over the control arms of these studies, whereas a revised risk of relapse for women who responded to treatment was calculated by multiplying the revised risk of relapse for those improving under usual care by the estimate of the relative risk of relapse given in the original meta-analysis.

93

Resource utilisation and cost inputs

The estimates of unit costs adopted by NICE were taken from Curtis and Netten;[232] these were updated by estimates taken from Curtis.[233] These were combined with the assumptions employed in the NICE guideline to estimate the total cost of each strategy reported in Appendix 6.

Utility weights

The utility weights used in the reconstructed NICE treatment model remained unchanged from the NICE guidance, with the exception that a beta-distribution was fitted to each parameter; the SE in each case was calculated from the respective standard distribution and sample size given in Revicki and Wood.[234]

Results of the treatment model

The probabilistic treatment model was run over 1000 simulations; the results are given in *Table 29*. In common with the NICE guideline model, listening home visits were found to be both more effective and more costly than structured psychological therapy; however, the respective ICER of £66,275 per additional QALY suggested that listening home visits were not a cost-effective alternative to structured psychological therapy based on conventional thresholds, representing a decision-maker's willingness to pay (WTP) for additional health outcome, applied to establish cost-effectiveness (typically in the region of £20,000–30,000 per QALY).[235] The ICER of £17,481 per QALY for structured psychological therapy versus usual care suggested that structured psychological therapy was a cost-effective treatment based on similar thresholds.

For each of the three WTP thresholds given, structured psychological therapy had a greater than 50% probability of being the cost-effective strategy; for a WTP threshold of £20,000 per QALY, listening home visits were only slightly more likely to be cost-effective than the strategy of usual care.

Implications of the treatment model

Structured psychological therapy was found to be a cost-effective treatment for women with PND. Although listening home visits were slightly more effective, the higher cost of this strategy resulted in an ICER of listening home visits versus structured psychological therapy that was well above conventional cost-effectiveness thresholds considered to represent value for money in the NHS. As such, in constructing the identification model it was assumed that structured psychological therapy would be provided as the first-line

treatment to all postnatal women identified with depression.

The identification model
Overview

The identification model took the form of a decision tree and comprised five components – a 'diagnostic' component and four mutually exclusive 'treatment' components – representing the pathways of care, costs and outcomes for (1) true-positive, (2) false-negative, (3) true-negative and (4) false-positive cases. The identification model was used to evaluate the costs and outcomes of a range of alternative formal identification strategies compared with current practice (i.e. opportunistic case finding identified as part of routine consultations with a GP and/or health visitor). It was assumed that each postnatal woman entered the diagnostic component of the model at 6 weeks postnatally, at which time a formal identification strategy may be employed to complement any diagnosis of PND by the woman's GP or health visitor. Depending upon the woman's actual state of depression and her diagnosis she was then assigned to one of the four treatment components and followed up for 52 weeks until the model end point at 58 weeks postnatally.

To simplify the detailed description of the model each of the five model components is discussed separately. Before this an overview of the strategies considered in the model is provided.

Strategies evaluated

As previously stated the decision problem should ideally include the comparison of all identification strategies that could feasibly be used in the NHS. However, in practice, these options may be constrained by the availability of evidence and the structural complexity of any model. The review of the validity of methods to identify PND, reported in Chapter 5, was used to inform the identification strategies considered in the economic analysis and provided the basis for selecting strategies.

Strategies were only considered in the economic analysis if there were sufficient data reported to be able to combine results and produce pooled summary estimates of sensitivity and specificity based on evaluations undertaken postnatally. In Chapter 5, studies were separately examined according to disease classification (major depression only, major or minor depression, any psychiatric disorder or other). The economic model focused on those classifications for which

TABLE 29 *Results of the reconstructed NICE treatment model*

Identification option	QALYs	Cost[a]	ICER	Prob. CE for max. WTP[b]		
				£20,000	£30,000	£40,000
Usual care	0.7036	£0.00	N/A	0.2200	0.0370	0.0060
Structured psychological therapy	0.7489	£792.10	£17,481	0.5040	0.5490	0.5360
Listening home visits	0.7513	£946.48	£66,275	0.2760	0.4140	0.4580

CE, cost-effective; ICER, incremental cost-effectiveness ratio; N/A, not applicable; QALYs, quality-adjusted life-years, WTP, willingness to pay.
a   All stated costs are incremental to those of usual care.
b   The probability that a strategy was more CE than all others given a particular maximum WTP for an additional QALY.

the data appeared most widely reported – major depression alone (DSM or equivalent) and major/minor depression. Given the wider range of instruments and cut points reported for major/minor depression, this classification (and associated diagnostic data) was used to inform the base-case economic analysis. This classification was also consistent with the population considered in the treatment model reported previously. Major depression alone was considered as part of the sensitivity analysis.

For the comparison of major or minor depression, results from the bivariate meta-analysis of sensitivity and specificity were reported for a range of instruments and cut points. Specifically, the bivariate analysis considered the EPDS (cut points 7–16), BDI (cut point 10) and HAMD (cut point 11). However, as a single study provided the data to pool for the HAMD (with an associated I^2 value of 0), the evidence for this strategy was not considered sufficiently robust for the purposes of the economic analysis. Therefore, the main strategies considered in the base-case analysis were the EPDS (cut points 7–16) and BDI (cut point 10) compared with current practice (i.e. routine case identification without the formal use of a diagnostic instrument). The different EPDS cut points were assessed as separate strategies (10 in total), which enabled an evaluation of the trade-off between the different sensitivity and specificity values to be considered in terms of cost-effectiveness. All case identification strategies in the base-case model were modelled as one-time screens, such that the readministration of the same instrument was not modelled.

In addition to the base-case analysis, separate scenarios were considered that explored a range of alternative approaches. These are discussed more fully in later sections. The alternative approaches considered: (1) alternative classifications (i.e.

considering major depression only); (2) alternative identification strategies (including a separate scenario including the Whooley questions and a scenario in which a separate confirmatory strategy was used employing a gold standard instrument; and (3) alternative costs associated with the management of false positives.

The diagnostic component
Model structure and key assumptions
The diagnostic component is shown in *Figure 35* in the form of a decision tree. All postnatal women (whether depressed or non-depressed) entered the diagnostic component of the model at 6 weeks postnatally. It was assumed that at this time those women who were depressed *may* have been identified via routine care as being depressed, whereas those women who were not depressed *were not* incorrectly identified as being depressed as part of routine care.

If no formal identification strategy was employed then depressed women positively identified via routine care entered the 'true-positive' treatment component of the model; depressed women not identified by routine care entered the 'false-negative' treatment component, whereas women who were non-depressed (and so not diagnosed) entered the 'true-negative' treatment component.

If a formal identification strategy such as the EPDS was employed then it was assumed that the EPDS was administered to *all* postnatal women at 6 weeks postnatally. Women were assumed to be diagnosed as depressed if they were positively identified by *either* the formal identification strategy *or* routine care. Hence, the value of identification strategies was determined not by the overall probability of identifying women with PND, but rather by the difference between this probability and the probability of identifying women with PND by

95
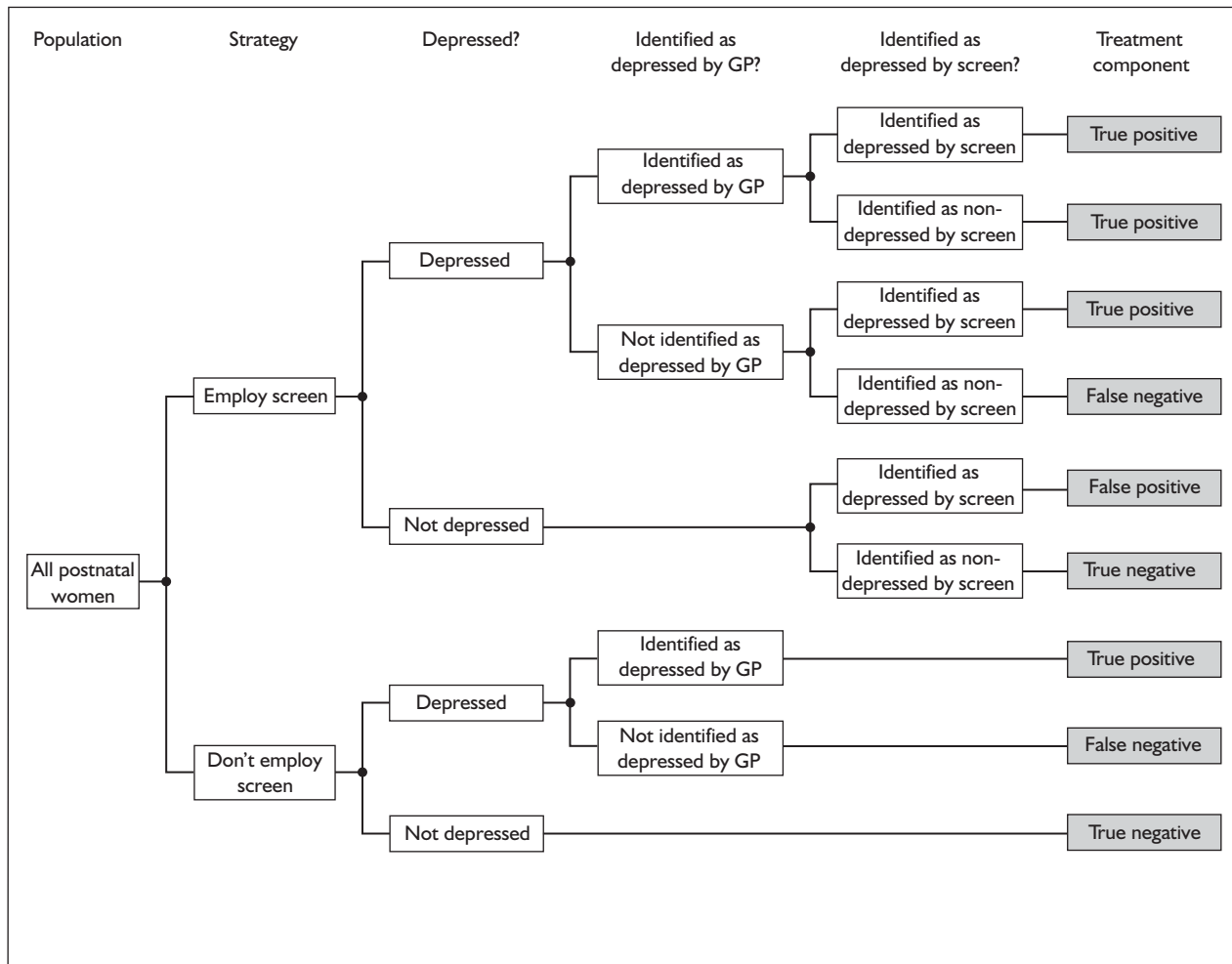
FIGURE 35 *The diagnostic component.*

routine care alone. Depressed women correctly diagnosed as depressed entered the 'true-positive' treatment component of the model; depressed women not diagnosed as depressed entered the 'false-negative' treatment component; non-depressed women not diagnosed as depressed entered the 'true-negative' treatment component; and non-depressed women wrongly diagnosed as depressed entered the 'false-positive' treatment component. Note that the terms 'true positive', 'false negative', etc. were used to relate a woman's *overall* diagnosis (including any identification via routine care) to her actual state of depression; the terms should not be conflated with the diagnostic accuracy results of the formal identification strategies.

An important structural assumption was that all postnatal women entered the relevant treatment component at 6 weeks postnatally (immediately following the identification strategy when given). It was therefore assumed that, in cases in which depression was diagnosed by routine care earlier in the postnatal period, no treatment was administered until 6 weeks postnatally and only then if the depression persisted at this time.

Parameter inputs
The parameter inputs incorporated into the diagnostic component are given in *Tables 30–32*. The estimate of the prevalence of major and minor depression among postnatal women at 6 weeks postnatally (11.3%) was taken from Gaynes *et al.*[27] for the end date '2 months PP'. It was assumed that the parameter was normally distributed, with the SE calculated from the given 95% CI (7.7% to 16.2%).

In the absence of estimates of the probability that PND was identified via routine care from a study reporting results in a PND population, the estimate of the probability that PND was detected via routine care at 6 weeks postnatally was derived from a study of subjects with depression or anxiety by Kessler *et al.*[236] This study found that 34 of 88 patients with depression or anxiety attending a general practice in Bristol in 1997 were detected by the GP at baseline, with 22 of the remaining

TABLE 30 *Input parameters – prevalence and routine case identification*

Input parameter	Mean	SE	Distribution	Source
Prevalence of major or minor depression	0.1130	0.0221	Normal	Gaynes *et al.*, 2005[27]
Probability of identification of depression at 6 weeks postnatally via routine care	0.3864	0.0516	Beta	Kessler *et al.*, 2002[236]
SE, standard error.				

54 detected by the GP during 3 years of follow-up (this latter finding was returned to in the sensitivity analysis). Taking the baseline to be 6 weeks postnatally, the probability that PND was detected by a GP at this time was assumed to take a beta (34,54) distribution with a mean of 34/88.

The sensitivity and specificity of the alternative identification strategies were derived from the bivariate meta-analysis reported in Chapter 5. For the purposes of the probabilistic model the sensitivity and specificity were modelled on the log-odds scale using normal distributions. The correlation between sensitivity and specificity, which was integral to the bivariate approach, was reflected in the probabilistic analysis using the Cholesky decomposition of the covariance matrix. The input distributions are reported in *Table 31*.

It was assumed that all postnatal women, whether depressed or non-depressed, made the same number of visits to a GP and received the same number of visits by a health visitor during the first 6 weeks postnatally. As such, any costs associated with these visits 'net out' of the analysis and so were not considered. The only costs relevant to the diagnostic component were those of administering the identification strategy (when applicable).

It was also assumed that all identification strategies were administered via a health visitor as part of a routine visit. As such, only the marginal cost of administration was considered (i.e. the additional time taken to administer an instrument over and above that required for a routine visit). The EPDS and BDI were assumed to require 5 additional minutes of the health visitor's time. In addition, the cost of the license fee of US$2 per test (£1.02 at current exchange rates in 2008) for the use of the BDI was included. The total costs of administrating each test are reported in *Table 32*.

The 'true-positive' treatment component
Model structure and key assumptions
The 'true-positive' treatment component is shown in *Figure 36* in the form of a decision tree. All

women who entered this treatment component were depressed and were correctly diagnosed as such by routine care and/or a formal identification strategy (when applicable).

The structure of this component was adapted from that of the reconstructed NICE treatment model discussed earlier. As it was assumed that all postnatal women diagnosed with depression were offered structured psychological therapy, the usual care and listening home visits arms of the reconstructed NICE treatment model were irrelevant and so were not considered. All other structural assumptions given in the reconstructed NICE treatment model remained.

Parameter inputs
The parameter inputs for this component were those used in the reconstructed NICE treatment model reported in Appendix 6.

The 'false-positive' treatment component
Model structure and key assumptions
The 'false-positive' treatment component is given in *Figure 37*. All women who entered this treatment component were non-depressed, but were wrongly diagnosed as having depression by the identification strategy. It was assumed in the diagnostic component that non-depressed women were not incorrectly diagnosed as depressed as part of routine care. Hence, the false-positive costs were only considered for the formal identification strategies. This could be considered a conservative approach towards evaluating the use of formal identification approaches as it is possible that some of the women identified as false positives via these approaches would also have been incorrectly diagnosed as part of routine care and, as such, some of the costs attributed to the identification strategies would have also been incurred as part of routine care. However, equally it could be argued that the process of routine case detection is potentially independent from the process of formal identification strategies and that the issue of false positives associated with routine care could affect all strategies in the same way.

TABLE 31 *Performance of diagnostic tests*

Diagnostic test	Sensitivity (major/minor depression)				Specificity (major/minor depression)					
	Log-odds	Variance	Distribution	Mean	Log-odds	Variance	Distribution	Mean	Co-variance	Source
EPDS (cut point 7)	2.3346	1.7439	Normal	0.9117	0.7075	0.4125	Normal	0.6699	–0.1164	Bivariate meta-analysis (Chapter 5)
EPDS (cut point 8)	2.3379	0.9962	Normal	0.9120	1.0744	0.7713	Normal	0.7454	0.1702	
EPDS (cut point 9)	1.7565	1.0652	Normal	0.8528	1.5274	0.8888	Normal	0.8216	0.0457	
EPDS (cut point 10)	1.4962	0.3514	Normal	0.8170	1.8373	1.3740	Normal	0.8626	–0.0524	
EPDS (cut point 11)	0.9548	0.5324	Normal	0.7221	2.3261	1.6619	Normal	0.9110	–0.3808	
EPDS (cut point 12)	0.7563	0.2095	Normal	0.6805	2.5959	1.4008	Normal	0.9306	–0.2222	
EPDS (cut point 13)	0.6716	0.6313	Normal	0.6619	2.5431	0.7443	Normal	0.9271	–0.3553	
EPDS (cut point 14)	0.1326	0.0008	Normal	0.5331	3.1069	1.1104	Normal	0.9572	–0.0282	
EPDS (cut point 15)	–0.4431	0.0301	Normal	0.3910	3.7755	0.2873	Normal	0.9776	0.0930	
EPDS (cut point 16)	–0.7875	0.3989	Normal	0.3127	4.5233	0.1812	Normal	0.9893	0.2689	
BDI (cut point 10)	0.9408	0.4659	Normal	0.7193	2.2675	0.4583	Normal	0.9061	0.4621	

Diagnostic test	Sensitivity (major depression only)				Specificity (major depression only)					
	Log-odds	Variance	Distribution	Mean	Log-odds	Variance	Distribution	Mean	Co-variance	Source
EPDS (cut point 7)	3.1781	0.0000	Normal	0.9600	–0.1879	0.7037	Normal	0.4532	–0.0001	Bivariate meta-analysis (Chapter 5)
EPDS (cut point 8)	2.7487	0.0263	Normal	0.9398	0.3168	0.4941	Normal	0.5786	0.1139	
EPDS (cut point 9)	2.4998	0.0574	Normal	0.9241	0.6582	0.4766	Normal	0.6589	0.1653	
EPDS (cut point 10)	2.4026	0.1657	Normal	0.9170	1.2129	0.4529	Normal	0.7708	0.2740	
EPDS (cut point 11)	1.9174	0.1906	Normal	0.8719	1.6353	0.6600	Normal	0.8369	0.1095	
EPDS (cut point 12)	1.7886	0.0580	Normal	0.8568	1.9138	0.9112	Normal	0.8714	–0.2299	
EPDS (cut point 13)	1.3385	0.0040	Normal	0.7922	2.1361	0.5333	Normal	0.8944	–0.0395	
EPDS (cut point 14)	0.9766	0.0301	Normal	0.7264	2.4369	0.6733	Normal	0.9196	0.0692	
EPDS (cut point 15)	0.6283	0.0350	Normal	0.6521	3.0953	0.6562	Normal	0.9567	–0.0963	
EPDS (cut point 16)	0.3863	0.0110	Normal	0.5954	3.6035	1.1603	Normal	0.9735	–0.1130	

BDI, Beck Depression Inventory; EPDS, Edinburgh Postnatal Depression.

TABLE 32 *Cost inputs*

Cost element	Value	Distribution	Source
BDI license fee (per test)	US$2.00	Fixed	www.harcourtassessment.com (accessed 1 April 2008)
Health visitor (per hour)	£91.00	Fixed	Curtis, 2007[233]
Exchange rate	**Value**	**Distribution**	**Source**
US$ to UK£	£0.51	Fixed	www.FT.com (accessed 1 April 2008)
Identification strategy	**Resources used (per test)**		**Source**
EPDS	Health visitor (5 minutes)		Cox *et al.*, 1987[16]
BDI	Health visitor (5 minutes) License fee		www.harcourtassessment.com (accessed 1 April 2008)
Total cost (per test)	**Value**	**Distribution**	**Source**
EPDS	£7.57	Fixed	Estimated from above inputs
BDI	£8.59	Fixed	
BDI, Beck Depression Inventory; EPDS, Edinburgh Postnatal Depression.			
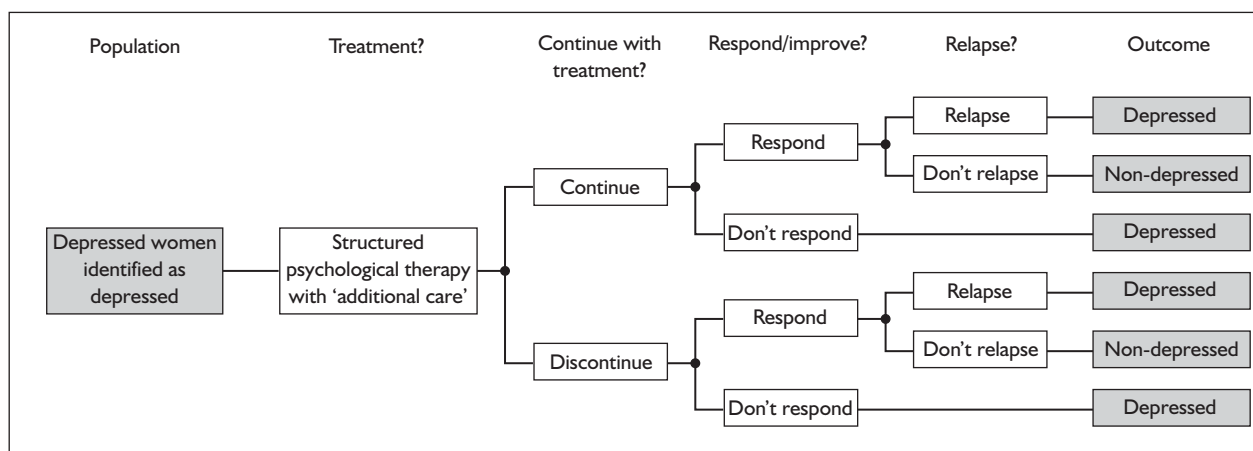
FIGURE 36 *The 'true-positive' treatment component.*
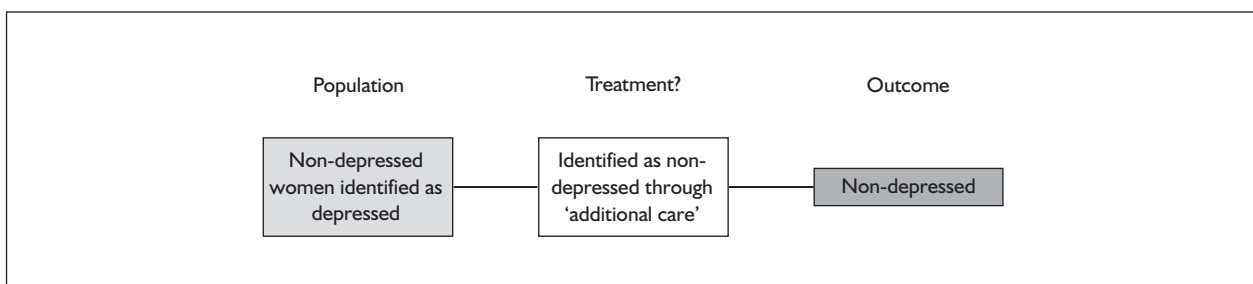
FIGURE 37 *The 'false-positive' treatment component.*

In common with the 'true-positive' component it was assumed that all women were initially offered structured psychological therapy with the preceding 'additional care'. However, it was assumed that the true non-depressed state of each woman would become apparent during the 'additional care' phase and as a result structured psychological therapy itself would not be provided. Those women wrongly diagnosed as depressed by the screen therefore incurred only the cost of 'additional care' (this assumption was returned to in the sensitivity analysis) and not the costs of psychological therapy. Furthermore, it was assumed that incorrectly identifying a non-depressed woman as depressed carried a loss in quality of life (i.e. no decrement in utility was assigned in the model) and that all women non-depressed at 6 weeks postnatally remained non-depressed until the model end point; as such, all women in this component experienced the HRQoL associated with remission throughout the follow-up period. Although in reality women who were not suffering PND at 6 weeks postnatally may become depressed at some point in the following year, the evaluation of identification strategies was based on the 6-week time point and hence has no impact on the subsequent management of women who were non-depressed at 6 weeks (aside from the costs of false positives) during the remaining course of the model. Consequently, this assumption had no impact on the incremental cost-effectiveness estimates reported here.

Parameter inputs
The only parameter inputs for this component of the model were the cost of additional care and the utility value for the state of remission (see Appendix 6).

The 'false-negative' treatment component
Model structure and key assumptions
The 'false-negative' treatment component is given in *Figure 38*. All women who entered this treatment component were depressed, but had not been identified either by routine care or by the formal identification strategies.

As the depression remained unidentified, each woman initially received usual care for non-depressed postnatal women. As in the reconstructed NICE treatment model there was a possibility that the woman's depression might improve under usual care by 14 weeks postnatally (8 weeks after the model start point), in which case the woman either remained in a state of remission or relapsed over the remaining 44-week follow-up period.

If the woman did not improve under usual care then it was assumed that she would make an additional visit to her GP – this was assumed to take place half-way through the follow-up period, at 32 weeks postnatally. At this visit there was a possibility that the woman's depression would be identified by routine care and structured psychological therapy immediately offered as treatment. The structure of the model following this node was adapted from the 'true-positive' treatment component: if the woman responded to treatment she was assumed to have experienced a linear improvement in HRQoL from 32 weeks postnatally to 40 weeks postnatally; if she did not respond to treatment she remained depressed until the model end point at 58 weeks postnatally. If the woman responded but then relapsed it was assumed that she relapsed at the same rate as those women who relapsed following treatment initiated at 6 weeks postnatally (i.e. a linear deterioration in HRQoL over 44 weeks); as the model end point was 18 weeks into this 44-week linear deterioration, those who relapsed after responding to treatment initiated half-way through the follow-up period were assumed to be in a state of 'partial relapse' at the model end point. The state of 'partial relapse' was used simply to refer to the calculations made in relation to the HRQoL inputs as opposed to representing a separate health state of the model. Meanwhile, if the woman responded to treatment, but did not relapse then she was assumed to remain in a state of remission until the model end point, while if the woman's depression was not identified by routine care at the additional GP visit it was assumed that she continued receiving usual care and remained depressed until the model end point.

Parameter inputs
The parameter inputs for this component were those used in the reconstructed NICE treatment model (Appendix 6) and the probability of GP identification of depression at 32 weeks postnatally. The estimate of the probability that PND was detected by a woman's GP at 32 weeks postnatally was taken from Kessler *et al.*[236] This study found that 22 of the remaining 54 patients with depression who were not identified at baseline were eventually detected by the GP during 3 years of follow-up. Because of the lack of more suitable data, this finding was used to approximate the probability of GP identification of depression at the model midpoint (32 weeks postnatally); the probability that PND was detected by a GP at this time was assumed to take a beta (22,32) distribution with a mean of 22/54 (*Table 32*).
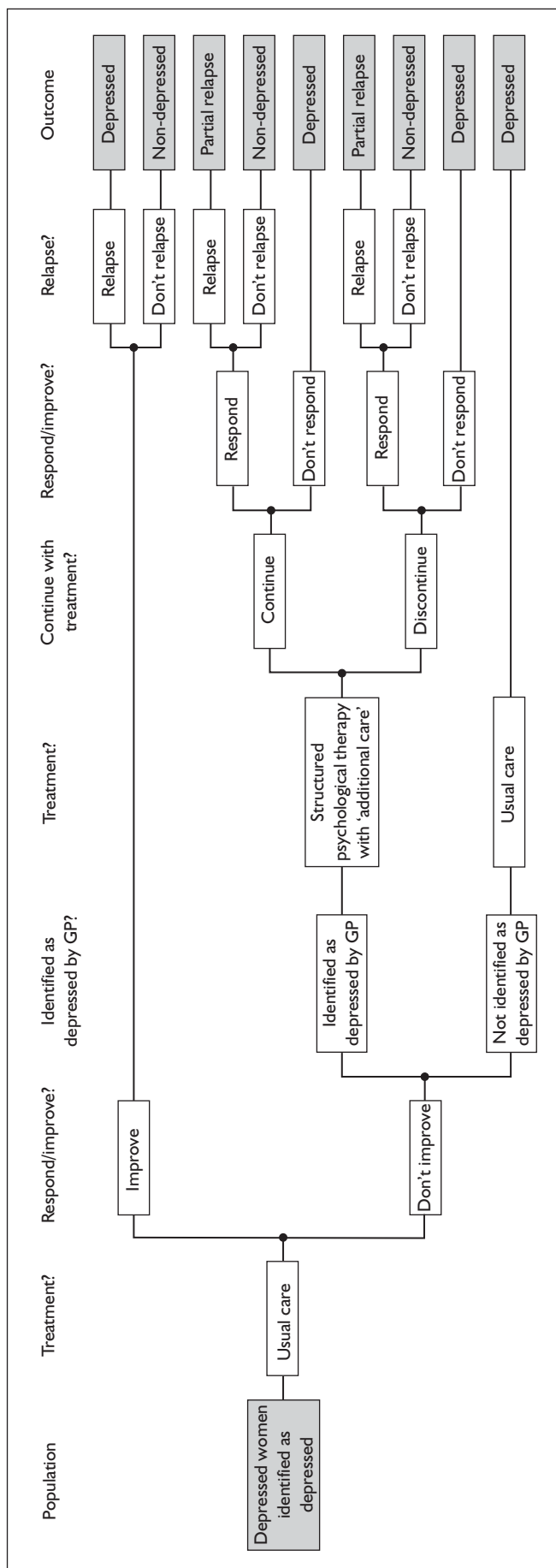
FIGURE 38 *The 'false-negative' treatment component.*

TABLE 33 *Probability of GP identification of depression at 32 weeks postnatally*

	Mean	SE	Distribution	Source
Probability of GP identification of depression at 32 weeks postnatally	0.4074	0.0663	Beta	Kessler *et al.,* 2002[236]

SE, standard error.

The 'true-negative' treatment component
Model structure and key assumptions

The 'true-negative' treatment component is given in *Figure 39*. All women who entered this treatment component were non-depressed and had either not been diagnosed at all or been correctly identified as non-depressed by routine care and/or a formal identification approach (when applicable).

Each woman received usual care for non-depressed postnatal women. As the model only considered the additional costs related to the management of PND, the costs of usual care incurred by women who were not depressed were not included in the model. It was assumed that all women non-depressed at 6 weeks postnatally would remain non-depressed until the model end point; as such, all women in this component experienced the HRQoL associated with remission throughout the follow-up period.

Parameter inputs

The only parameter input for this component of the model was the utility value for the state of remission (Appendix 6).

Analytic methods of the identification model

The model was developed in Microsoft EXCEL.[237] The Monte Carlo simulation was run for 10,000 iterations. The model was run several times representing the base-case analysis and alternative scenarios considered as part of the sensitivity analysis.

The results of the model are presented in two ways. First, mean costs and QALYs of the alternative identification strategies are presented and their cost-effectiveness compared, estimating ICERs as appropriate, using standard decision rules.[238] The ICERs examined the additional costs that one strategy incurred over another and compared this with the additional benefits. When more than two strategies were being compared the ICERs were calculated using the following process:

- The identification strategies were ranked in terms of cost (from the least expensive to the most costly).

- If a strategy was more expensive and less effective than the previous strategy then that strategy was said to be dominated and was excluded from the calculation of the ICERs.
- The ICERs were calculated for each successive alternative, from the least expensive cost to the most costly. If the ICER for a given strategy was higher than that of the next most effective strategy then that strategy was ruled out on the basis of extended dominance.
- Finally, the ICERs were recalculated excluding any strategies that were ruled out using the notions of dominance and extended dominance.

The advantage of entering input parameters as uncertain variables in the probabilistic analysis was that the uncertainty could be propagated through the model and reflected in model outputs representing uncertainty surrounding the decision itself. To present the uncertainty in the cost-effectiveness of the alternative strategies, cost-effectiveness acceptability curves (CEACs) were calculated.[239,240] These figures show the probability that each strategy was more cost-effective than the other three using alternative values for the maximum value that the health service was willing to pay for an additional QALY in these patients.

Although the CEACs provide a useful graphical representation of the uncertainty associated with the probability that individual strategies were cost-effective over a range of threshold values, the results of the CEACs could only be used to identify the optimal implementation decision under a restrictive set of assumptions. This was because the strategy with the highest probability of being cost-effective did not necessarily have the highest expected pay-off (i.e. net benefit), and would only do so when the distribution of these pay-offs was symmetrical. This limitation could be overcome by using a cost-effectiveness frontier to indicate which strategy was optimal (and the associated probability that this strategy was the most cost-effective) across the range of values representing the maximum amount that the NHS was WTP for an additional QALY.[239]
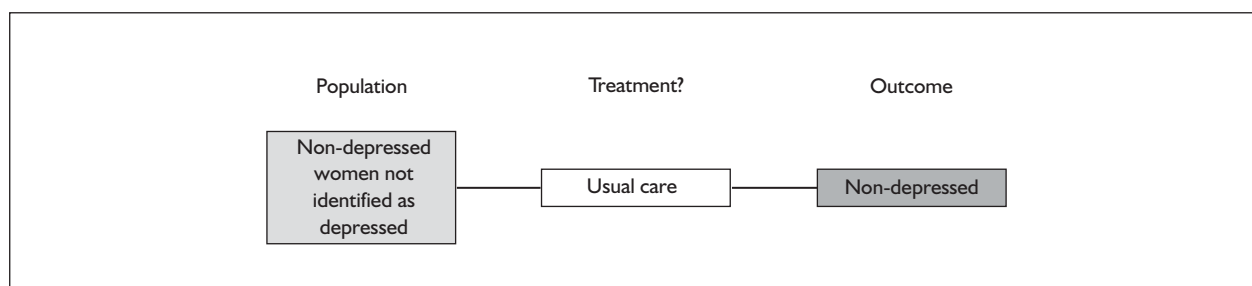
FIGURE 39 The 'true-negative' treatment component.

Results

The results are presented in two stages: those of the base-case analysis, in which the assumptions set out in previous sections were employed; and those of the sensitivity analysis, in which the impact of employing alternative assumptions to those of the base case was explored.

Base-case results

A summary of the results of the base-case analysis is given in *Table 34*. Routine care was the least costly and least effective strategy. When the identification strategies were ranked in terms of cost (from the least expensive to the most expensive) the EPDS cut points 7 and 13 and the BDI cut point 10 were more expensive and less effective than the previous strategies in the list and thus were ruled out on the grounds of dominance. After calculating the

ICERs for the non-dominated strategies, the ICER for the EPDS cut point 15 was found to be higher than that of the next more effective strategy on the ranked list and thus was ruled out on the grounds of extended dominance. Of the remaining non-dominated identification strategies, the EPDS at a cut point of 16 was the next most costly and effective strategy compared to routine care, with an associated ICER of £41,103 per QALY. The ICER of the EPDS at a cut point of 14 was £49,928 per QALY compared to the EPDS cut point 16. The ICER of the EPDS at lower cut points (e.g. 8, 9–11) exceeded £100,000 per QALY.

In general, the ranking of identification strategies appeared to be driven by their specificity, such that strategies with a high specificity (e.g. EPDS cut point 16) were associated with more favourable ICERs than strategies with a lower specificity (but correspondingly a higher sensitivity), suggesting

TABLE 34 Base-case results

Identification option	QALY	Cost	ICER[a]	Prob. CE for max. WTP[b]		
				£20,000	£30,000	£40,000
Routine care	0.846	£49.29	N/A	0.8765	0.5869	0.393
EPDS 16	0.846	£73.49	£41,103	0.0221	0.0614	0.0684
EPDS 15	0.846	£80.95	ED	0.007	0.0182	0.0198
EPDS 14	0.847	£94.21	£49,928	0.0158	0.0439	0.0527
EPDS 13	0.847	£110.47	D	0.0052	0.0253	0.0425
EPDS 12	0.847	£109.95	£56,697	0.0177	0.0611	0.0877
BDI 10	0.847	£121.51	D	0.0115	0.0507	0.0895
EPDS 11	0.847	£118.82	£113,411	0.0186	0.0587	0.0853
EPDS 10	0.847	£140.44	£120,968	0.0172	0.0564	0.089
EPDS 9	0.847	£156.95	£245,210	0.0065	0.026	0.0464
EPDS 7	0.847	£215.07	D	0.0001	0.0004	0.0012
EPDS 8	0.847	£187.32	£272,463	0.0018	0.011	0.0245

BDI, Beck Depression Inventory; CE, cost-effective; EPDS, Edinburgh Postnatal Depression Scale; ICER, incremental cost-effectiveness ratio; N/A, not applicable; QALY, quality-adjusted life-year; WTP, willingness to pay.
a   D represents a strategy ruled out through dominance; ED represents a strategy ruled out through extended dominance.
b   The probability that a strategy is more CE than all others given a particular maximum WTP for an additional QALY.

that the costs of managing false positives represents a key driver of these results.

At each of the three WTP thresholds considered, the strategy with the highest individual probability of being cost-effective was routine case detection. However, at a threshold of £30,000 per QALY there was a probability of 0.4131 (i.e. 41%) that routine care was less cost-effective than a policy of using formal identification approaches. However, despite the formal identification strategies having a combined probability of being cost-effective which exceeded that of routine care, the probabilities associated with each of the individual strategies were low. This suggested that, although formal identification approaches had a higher chance of being cost-effective than routine care, there was significant uncertainty between the separate strategies about which should be the optimal strategy based on cost-effectiveness considerations.

The base-case cost-effectiveness frontier is given in *Figure 40* for WTP threshold values up to £100,000 per QALY, demonstrating which strategy was cost-effective (and the probability that this was the most cost-effective) across this range. The high uncertainty surrounding the decision between individual identification strategies was clearly evident. For threshold values above £41,103 per QALY, the optimal identification strategy across their respective ranges had a very low probability of being cost-effective.

Sensitivity analysis

The sensitivity analysis explored the impact of considering alternatives to four key assumptions of the base-case analysis. Four separate scenarios were considered:

- *Scenario 1* – In the base-case analysis the total cost incurred by a non-depressed woman incorrectly identified as depressed by the identification strategy (i.e. the cost of a false positive) was assumed to be £414 – the full cost of 'additional care' (consisting of one community psychiatric nurse visit of 1 hour, three GP visits of 10 minutes each and four health visitor home visits of 45 minutes each) considered in the treatment model. The results of the base-case analysis showed that the ranking of identification strategies in cost-effectiveness terms appeared to largely determined by their specificity, such that strategies with the highest specificity (e.g. EPDS cut point 16) were associated with more favourable cost-effectiveness estimates and the ICERs of strategies with lower specificity (but higher sensitivity) were markedly higher and exceeded conventional thresholds used to establish value for money in the NHS. The robustness of the results to alternative assumptions concerning the management and costs of false-positive diagnoses was explored using separate scenarios. An alternative assumption was employed whereby it was assumed that false positives would be correctly diagnosed with a single GP consultation (as opposed to the complete 'package' of additional care assumed in the base-case analysis) and that no further costs beyond this would be incurred for this group. Hence, the impact on the results was explored when the total cost of such false-positive diagnoses was £25.50 – the cost of a single 10-minute
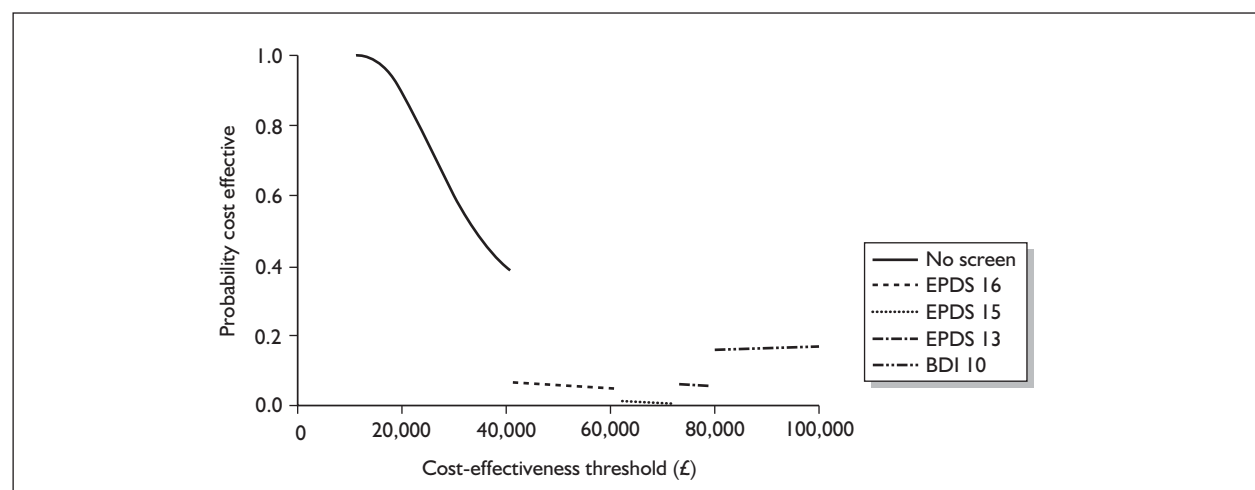
FIGURE 40 *Base-case cost-effectiveness frontier.*

GP appointment. An alternative 'best-case' scenario for the identification strategies was also considered in which the total cost of a false-positive diagnosis was assumed to be zero, highlighting the most optimistic estimate that could be assumed for the formal identification strategies.

- *Scenario 2* – The base-case approach assumed that only a single identification strategy was carried out at 6 weeks postnatally. This approach was employed because of the absence of robust data (reported in Chapter 5) on the test performance of repeating single identification strategies or using combinations of individual strategies. However, because of the importance of managing false positives identified as part of the base-case analysis, the impact of using a gold standard identification strategy (such as SCID) in patients positively identified using the alternative identification strategies was evaluated. By applying a gold standard reference strategy as part of a confirmatory approach to managing positive cases identified by individual identification strategies it was assumed that any false positives would be correctly diagnosed by the confirmatory identification approach, thus incurring only the additional costs of the confirmatory screen as opposed to the full costs of additional care considered in the base-case analysis. However, in contrast to the approach employed in the first set of scenarios (which evaluated alternative costs for managing false positives only), the cost of the confirmatory screen was applied to all patients who tested positive in the initial identification strategy (i.e. both true and false positives). The impact of using SCID as a confirmatory strategy on those women identified as depressed by the identification strategy was considered, with the cost-effectiveness of such 'combined' strategies compared with the cost-effectiveness of existing single identification strategies and the cost-effectiveness of the strategy of routine care.
- *Scenario 3* – The base case considered only the EPDS and the BDI as possible identification strategies. In this scenario the impact of considering the Whooley questions (with the third confirmation question) as a possible alternative identification strategy was explored. The Whooley questions were not considered in the base case because of the lack of data available to pool estimates as part of the bivariate meta-analysis and the concerns noted in Chapter 5 over the absence of data in a postnatal population. However, as part of the

sensitivity analysis it was considered important to explore the robustness of the base case to the inclusion of this strategy to directly address current policy and practice in the UK.
- *Scenario 4* – The base case considered those women defined as having major or minor depression. In this scenario the impact of considering only those women defined as having major depression was explored.

Scenario 1: The total cost of false-positive diagnoses

The impact of assuming lower costs associated with false-positive diagnoses had two important consequences: (1) the ICERs of formal identification approaches became more favourable compared with the base-case analysis, with estimates for particular strategies within conventional thresholds considered to represent value for money in the NHS; and (2) the ranking of (non-dominated) identification strategies altered, such that the ranking of treatment appeared less dependent upon the specificities of the individual strategies. Hence, the balance between the benefits of true positives and the costs of false positives became more evident in these scenarios.

Although routine care remained the least costly and least effective strategy when a false-positive cost of £25.50 (i.e. a single GP consultation) was assumed, the next more costly and effective strategy that was not ruled out on dominance grounds was the EPDS cut point 10 (as opposed to cut point 16 in the base-case analysis) (*Table 35*). The ICER associated with the EPDS cut point 10 compared with routine care was £29,186 per QALY. This strategy extendedly dominated those EPDS identification strategies with higher cut points and the BDI with a cut point of 10. The most cost-effective identification strategy was the EPDS cut point 8 at thresholds above £35,390 (EPDS cut point 7 was dominated by EPDS cut point 8). Although the strategy of routine care detection was the most likely to be cost-effective at each of the three WTP thresholds considered, at the two higher thresholds the probability that one of the identification strategies was cost-effective was 56.41% and 72.47% respectively.

When the cost of false-positive diagnoses was reduced to zero, the EPDS with a cut point of 8 dominated or extendedly dominated all other identification strategies and was cost-effective at a comparatively low WTP threshold of £25,980 per QALY (*Table 36*).

TABLE 35 *Results of the sensitivity analysis: cost of false-positive diagnoses £25.50*

Identification option	QALY	Cost	ICER[a]	Prob. CE for max. WTP[b]		
				£20,000	£30,000	£40,000
Routine care	0.8452	£50.33	N/A	0.7466	0.4359	0.2753
EPDS 16	0.8459	£71.77	ED	0.0008	0.0011	0.0006
EPDS 15	0.8460	£74.68	ED	0	0	0
EPDS 14	0.8463	£81.44	ED	0.0001	0	0
EPDS 13	0.8465	£86.72	ED	0.0114	0.0244	0.0241
EPDS 12	0.8465	£88.21	ED	0.0094	0.0139	0.0127
EPDS 11	0.8466	£90.08	ED	0.0236	0.0408	0.0425
BDI 10	0.8466	£90.39	ED	0.0287	0.0536	0.0599
EPDS 10	0.8468	£95.26	£29,186	0.0502	0.0934	0.1082
EPDS 9	0.8468	£96.27	ED	0.0558	0.1282	0.1651
EPDS 7	0.8469	£101.01	D	0.0156	0.0592	0.1015
EPDS 8	0.8469	£100.74	£35,390	0.0578	0.1495	0.2101

BDI, Beck Depression Inventory; CE, cost-effective; EPDS, Edinburgh Postnatal Depression Scale; ICER, incremental cost-effectiveness ratio; N/A, not applicable; QALY, quality-adjusted life-year; WTP, willingness to pay.
a   D represents a strategy ruled out through dominance; ED represents a strategy ruled out through extended dominance.
b   The probability that a strategy is more CE than all others given a particular maximum WTP for an additional QALY.

TABLE 36 *Results of the sensitivity analysis: cost of false-positive diagnoses zero*

Identification option	QALY	Cost	ICER[a]	Prob. CE for max. WTP[b]		
				£20,000	£30,000	£40,000
Routine care	0.8453	£50.05	N/A	0.7011	0.3912	0.2536
EPDS 16	0.8459	£71.02	ED	0	0	0
EPDS 15	0.8461	£73.77	ED	0	0	0
EPDS 14	0.8463	£79.59	ED	0	0	0
EPDS 13	0.8466	£84.13	ED	0.0023	0.0053	0.0073
EPDS 12	0.8466	£85.37	ED	0.0004	0.0009	0.0010
EPDS 11	0.8467	£86.41	ED	0.0052	0.0113	0.0138
BDI 10	0.8467	£87.53	ED	0.0008	0.0031	0.0052
EPDS 10	0.8469	£90.66	ED	0.0150	0.0290	0.0351
EPDS 9	0.8469	£91.11	ED	0.0501	0.1034	0.1287
EPDS 7	0.8470	£93.10	ED	0.1220	0.2436	0.2925
EPDS 8	0.8470	£93.86	£25,980	0.1031	0.2122	0.2628

BDI, Beck Depression Inventory; CE, cost-effective; EPDS, Edinburgh Postnatal Depression Scale; ICER, incremental cost-effectiveness ratio; N/A, not applicable; QALY, quality-adjusted life-year; WTP, willingness to pay.
a   ED represents a strategy ruled out through extended dominance.
b   The probability that a strategy is more CE than all others given a particular maximum WTP for an additional QALY.

Scenario 2: Employing SCID as a confirmatory strategy

The base case assumed that only a single identification strategy was employed at 6 weeks postnatally. An alternative was to use a confirmatory strategy such as SCID on those women initially identified as depressed. It was assumed that SCID had a sensitivity and specificity of 100% and took 30 minutes of a health visitor's time to administer (and so cost £45.50). The cost-effectiveness of 'combined' strategies (consisting of an identification strategy followed by a confirmatory strategy using SCID) was compared with the cost-effectiveness of the existing single identification strategies and with the cost-effectiveness of the strategy of routine case

detection. As such, there were 25 possible strategies (EPDS cut points 7–16, EPDS cut points 7–16 followed by confirmatory SCID, BDI cut point 10, BDI cut point 10 followed by confirmatory SCID, and routine case detection). The results of this scenario are reported in *Table 37*.

For all strategy types and cut points, adopting the combined strategy with SCID as a confirmatory strategy dominated the corresponding single identification strategies. No combined strategy was cost-effective for WTP thresholds below £33,776 per QALY and, as in the base case, the optimal EPDS cut point decreased as the WTP threshold increased. However, whereas in the base case the EPDS cut point 8 was cost-effective only for

TABLE 37 Results of the sensitivity analysis: considering SCID as a confirmatory strategy

Identification option	QALY	Cost	ICER[a]	Prob. CE for max. WTP[b]		
				£20,000	£30,000	£40,000
Routine care	0.8450	£50.11	N/A	0.8362	0.5261	0.3358
EPDS 16	0.8456	£75.40	D	0.0051	0.0076	0.0056
EPDS 16 + SCID	0.8456	£73.24	ED	0	0	0
EPDS 15	0.8457	£83.27	D	0.0001	0.0002	0
EPDS 15 + SCID	0.8457	£76.87	ED	0	0	0
EPDS 14	0.8460	£103.30	D	0.0010	0.0009	0.0003
EPDS 14 + SCID	0.8460	£85.01	ED	0.0006	0.0004	0.0005
EPDS 13	0.8462	£118.41	D	0.0002	0.0002	0.0002
EPDS 13 + SCID	0.8462	£91.30	£33,776	0.0114	0.0316	0.0377
EPDS 12	0.8462	£125.55	D	0.0035	0.0037	0.0042
EPDS 12 + SCID	0.8462	£93.33	ED	0.0117	0.0241	0.0227
BDI 10	0.8463	£128.04	D	0.0002	0.0014	0.0017
BDI 10 + SCID	0.8463	£95.62	ED	0.0193	0.0601	0.0770
EPDS 11	0.8463	£137.88	D	0.0039	0.0061	0.0053
EPDS 11 + SCID	0.8463	£95.87	ED	0.0186	0.0405	0.0485
EPDS 10	0.8465	£160.01	D	0.0036	0.0075	0.0085
EPDS 10 + SCID	0.8465	£102.49	£37,391	0.0296	0.0829	0.1075
EPDS 9	0.8465	£169.77	D	0.0004	0.0007	0.0008
EPDS 9 + SCID	0.8465	£104.04	ED	0.0282	0.0943	0.1396
EPDS 7	0.8466	£218.45	D	0	0	0
EPDS 7 + SCID	0.8466	£111.32	D	0.0036	0.0256	0.0571
EPDS 8	0.8466	£198.06	D	0.0002	0.0003	0.0004
EPDS 8 + SCID	0.8466	£110.00	£50,408	0.0226	0.0858	0.1466

BDI, Beck Depression Inventory; CE, cost-effective; DSM, *Diagnostic and Statistical Manual of Mental Disorders*; EPDS, Edinburgh Postnatal Depression Scale; ICER, incremental cost-effectiveness ratio; N/A, not applicable; QALY, quality-adjusted life-year; SCID, Structured Clinical Interview for DSM; WTP, willingness to pay.
a D represents a strategy ruled out through dominance; ED represents a strategy ruled out through extended dominance.
b The probability that a strategy is more CE than all others given a particular maximum WTP for an additional QALY.

thresholds above £272,463 per QALY, the EPDS cut point 8 with confirmatory SCID was cost-effective for all thresholds above £50,408 per QALY.

Scenario 3: Considering the Whooley questions as an alternative identification strategy

The base case considered only the EPDS and the BDI as possible identification strategies based on the robustness of the data available (reported in Chapter 5). The impact of considering the Whooley questions (with the third confirmation question) as a possible alternative identification strategy was evaluated using a sensitivity of 0.96 (95% CI 0.86 to 0.99) and a specificity of 0.89 (95% CI 0.87 to 0.91) as reported in Arroll *et al.*[39]

In common with the base-case analysis the use of the EPDS was associated with the lowest ICER (£41,175 per QALY compared with routine care, *Table 38*). However, in contrast to the base-case analysis all other strategies, with the exception of the Whooley questions, were ruled out on the grounds of dominance or extended dominance. The Whooley strategy now dominated many of the EPDS strategies (e.g. at cut points 7–10). Although the use of the Whooley questions appeared more effective than the use of the EPDS cut point 16, the additional costs of this strategy (primarily driven by the lower specificity associated with the Whooley

questions) resulted in an ICER for the Whooley questions compared with the EPDS cut point 16 of £46,538 per QALY.

Scenario 4: Considering women with major depression only

The base case considered those women defined as having major or minor depression. The impact of considering only those women defined as having major depression was considered in this scenario. It was assumed that the prevalence of major depression was that given in Gaynes *et al.*[27] (mean 0.068, SE 0.02), and that those women in a state of major depression experienced the HRQoL for severe depression given in Revicki and Wood[234] (mean 0.30, SE 0.04). The performance of each of the identification strategies was recalculated based on the bivariate results reported for major depression alone (DSM or equivalent) in Chapter 5. Because of the lack of suitable data, BDI cut point 10 was excluded from this analysis; therefore, this scenario compared the EPDS (cut points 7–16) and routine care.

The ICER for the identification strategy EPDS cut point 16 compared with the strategy of routine case detection was £23,195 per QALY (*Table 39*), as opposed to £41,103 per QALY in the base case. Of the remaining (non-dominated) alternatives, the use of lower EPDS cut points was associated with

TABLE 38 Results of the sensitivity analysis: considering the Whooley questions as an alternative strategy

Identification option	QALY	Cost	ICERa	Prob. CE for max. WTPb		
				£20,000	£30,000	£40,000
Routine care	0.8455	£49.34	N/A	0.8852	0.5978	0.3927
EPDS 16	0.8461	£73.54	£41,175	0.0208	0.0515	0.0590
EPDS 15	0.8462	£81.00	ED	0.0056	0.0158	0.0159
EPDS 14	0.8465	£94.25	ED	0.0149	0.0379	0.0399
EPDS 13	0.8467	£110.51	D	0.0060	0.0235	0.0295
EPDS 12	0.8467	£109.99	ED	0.0170	0.0565	0.0766
BDI 10	0.8468	£121.55	D	0.0094	0.0456	0.0738
EPDS 11	0.8468	£118.87	ED	0.0171	0.0567	0.0748
EPDS 10	0.8470	£140.48	D	0.0143	0.0498	0.0778
EPDS 9	0.8471	£156.99	D	0.0054	0.0220	0.0392
EPDS 7	0.8472	£215.11	D	0.0001	0.0002	0.0006
EPDS 8	0.8472	£187.36	D	0.0015	0.0084	0.0173
Whooley questions	0.8473	£130.16	£46,538	0.0027	0.0343	0.1029

BDI, Beck Depression Inventory; CE, cost-effective; EPDS, Edinburgh Postnatal Depression Scale; ICER, incremental cost-effectiveness ratio; N/A, not applicable; QALY, quality-adjusted life-year; WTP, willingness to pay.
a D represents a strategy ruled out through dominance; ED represents a strategy ruled out through extended dominance.
b The probability that a strategy is more CE than all others given a particular maximum WTP for an additional QALY.

TABLE 39 *Results of the sensitivity analysis: considering major depression only*

Identification option	QALY	Cost	ICER[a]	Prob. CE for max. WTP[b]		
				£20,000	£30,000	£40,000
Routine care	0.8376	£30.98	N/A	0.4321	0.2424	0.1660
EPDS 16	0.8393	£70.29	£23,195	0.1838	0.1661	0.1288
EPDS 15	0.8395	£76.93	£42,195	0.1179	0.1281	0.1125
EPDS 14	0.8397	£96.12	ED	0.0857	0.1271	0.1356
EPDS 13	0.8399	£107.36	£75,321	0.0473	0.0798	0.1003
EPDS 12	0.8400	£123.40	£90,930	0.0719	0.1274	0.1608
EPDS 11	0.8401	£133.48	ED	0.0445	0.0841	0.1187
EPDS 10	0.8402	£157.87	£212,593	0.0137	0.0343	0.0562
EPDS 9	0.8402	£198.71	ED	0.0016	0.0063	0.0126
EPDS 8	0.8403	£227.73	ED	0.0009	0.0026	0.0053
EPDS 7	0.8403	£271.07	£814,623	0.0006	0.0018	0.0032

BDI, Beck Depression Inventory; CE, cost-effective; EPDS, Edinburgh Postnatal Depression Scale; ICER, incremental cost-effectiveness ratio; N/A, not applicable; QALY, quality-adjusted life-year; WTP, willingness to pay.
a ED represents a strategy ruled out through extended dominance.
b The probability that a strategy is more CE than all others given a particular maximum WTP for an additional QALY.

increasing ICER estimates such that the ICER of the EPDS cut point 15 versus the EPDS cut point 16 was £42,195 per QALY and the ICER of the EPDS cut point 7 versus the EPDS cut point 10 was £814,623 per QALY.

Discussion

The results of the base-case analysis suggested that the use of formal identification strategies did not appear to represent value for money based on conventional thresholds of cost-effectiveness used in the NHS. However, the scenarios considered demonstrated that this conclusion was primarily driven by the costs of false positives assumed in the base-case model. Alternative assumptions employed in separate scenarios resulted in more favourable estimates of cost-effectiveness, such that the use of the EPDS as an identification approach for women with PND fell within these conventional thresholds.

It should be recognised that the costs of additional care assumed for the management of false positives (£414) represented a significant additional cost associated with the identification strategies that in reality remains highly uncertain. In the absence of reliable data on this aspect, a conservative approach was employed as part of the base-case assumptions. However, alternative and plausible

estimates (e.g. assuming a single GP appointment) or approaches (e.g. the use of a confirmatory SCID assessment) to managing false positives resulted in markedly more favourable cost-effectiveness estimates for formal identification approaches.

In addition to being a key driver in relation to the overall cost-effectiveness estimates, the costs of false-positive diagnoses were also central in terms of the relative cost-effectiveness of the different identification strategies. When the cost of a false-positive diagnosis was relatively high, as in the base case, the specificity of an identification strategy was an important contributor to its cost-effectiveness – in the base-case analysis the cost-effectiveness of the EPDS strategies with low cut points was less favourable because of their associated lower specificity. As the cost of a false-positive diagnosis fell, however, specificity became less important relative to sensitivity and, in the extreme case that such false diagnoses carried no additional cost, the EPDS cut point 8 (the strategy with the highest sensitivity) emerged as the optimal strategy. When the cost of a false-positive diagnosis was assumed to be that of a single GP attendance, the EPDS cut point 10 emerged as the optimal strategy in terms of cost-effectiveness, corresponding closely with the results presented in Chapter 5 based on the sROC curves for the alternative approaches in which the trade-off between sensitivity and specificity was

considered (but not in terms of their related costs and outcomes).

Although there was an absence of reliable published data on the costs of false positives, it would appear reasonable to conclude that the actual estimate lies somewhere between the full cost of the additional care considered in the base-case analysis and that assumed for a single GP appointment. Hence, a definitive answer to the question as to whether formal identification strategies were cost-effective, and, if they were, which individual strategy was optimal in cost-effectiveness terms, clearly requires further more reliable evidence in relation to the costs of managing false positives. However, the results presented here suggested that in these scenarios the most cost-effective identification approaches are the EPDS at a cut point of 10 or higher. It should also be recognised that the cost of false positives associated with routine care were not considered as part of the base-case analysis and that some of the costs attributed to the formal identification strategies would have also been incurred during routine practice. As such, the base-case results should be considered as representing conservative estimates as to the potential value for money of formal identification approaches.

There were limited published data available for estimating other parameters in the model, namely: the probability that PND was identified via routine care at six weeks, the risk of relapse and the utility weights. As a result, the estimates used in the model were derived from studies of general depressed populations (i.e. non post-natal populations), which represents a serious limitation of the model. Furthermore, the EPDS was the only identification strategy where there was sufficient data at more than one cut point to be able to combine results and produce pooled summary estimates of sensitivity and specificity; as such, the performance of other identification strategies could not be assessed.

A further issue is the degree to which the QALY is an appropriate measure of health outcome. While the QALY is ubiquitous throughout the health economic evaluation literature, it has been argued that (as currently constructed) it is an insensitive measure of outcomes in mental health care.[241] In the absence of a suitable alternative, the QALY was adopted to ensure comparability between the interventions considered here and those outside the field of mental health; however, the potential insensitivity of the QALY in this context should be considered when interpreting the results.

Interestingly, the use of a combination identification strategy with confirmatory SCID dominated (i.e. were less costly and equally effective) the counterpart single identification strategy. However, using the base-case assumption related to the cost of a false-positive diagnosis (£414), the combination strategies still did not reach the conventional thresholds of cost-effectiveness. The finding that the confirmatory use of the SCID dominated the same individual strategy without a confirmatory screen was an interesting finding that was also closely related to the general issue of the cost of managing false positives. The use of the confirmatory SCID provided an alternative approach to reducing the costs associated with detecting false-positive diagnoses, obviating the costs of 'additional care' assumed elsewhere. Clearly the confirmatory SCID strategies would be dominated by the existing single identification strategies if the cost of a false positive was assumed to be the cost of a single GP attendance (as the administration cost of the SCID exceeded the cost of a GP attendance). As such, these estimates presented as part of this scenario analysis should be seen as the most optimistic estimates of the potential cost-effectiveness of using more definitive diagnostic instruments as a confirmatory approach to managing patients positively identified as depressed from an initial identification strategy such as the EPDS. There also remains an important issue of whether such a strategy would be feasible to implement in practice and whether health visitors could be trained to deliver the instrument and to interpret the subsequent data without the additional input of a more specialist practitioner. However, it was evident that alternative approaches that might be considered more appropriate, for example use of the SCID by a trained practitioner in a hospital setting, would be markedly more expensive than that considered here and hence it would appear unlikely that such an approach would be more cost-effective than the single identification approaches.

The results presented for major depression alone suggested that the use of formal identification strategies may be cost-effective as part of an approach to the management of major depression even in scenarios in which the cost of managing false positives was high. In the scenario considered, use of the EPDS as an identification strategy fell within conventional cost-effectiveness thresholds. This was primarily because of the improved test performance associated with the diagnosis of major depression alone (i.e. the probability of false positives was reduced compared with the base-case analysis evaluating major and minor depression

together) and the relatively higher consequences associated with not detecting true positives as part of routine care (i.e. in terms of loss of quality of life through not receiving appropriate treatment). However, it should also be recognised that a higher proportion of women with major depression may be identified via routine practice than the estimate applied in the model, which could offset (partially or wholly) the improved cost-effectiveness estimates identified as part of this scenario.

Finally, it should be noted that the model focused on the costs and outcomes associated with the mother herself. Because of a lack of reliable evidence, no account was taken of the potential impact of successful identification and subsequent management of PND on other family members or the infant. Clearly if identification strategies and subsequent management have an important effect on these aspects then the results presented here will represent highly conservative estimates of the potential value of identification approaches.

Despite the limitations and uncertainties noted, the results presented here represent the first

attempt to formally evaluate the potential cost-effectiveness of alternative identification strategies for the management of women with PND. Equally important is that this evaluation provided a systematic, integrated and transparent approach to synthesising the available evidence on the diagnostic performance of identification strategies (reported in Chapter 5) with the best available evidence relating to the subsequent management of and outcomes for women with PND reflecting current policy and practice in the NHS. Clearly there remain a number of uncertainties in relation to the assumptions that underpin the model and also in terms of the uncertainty characterised by the probability distributions assigned to the inputs. As new evidence emerges, the assumptions and inputs of this model could be updated and the results re-estimated on an ongoing basis. However, the current uncertainties related to parameter inputs could also be considered in terms of their impact on existing decision uncertainty and could be used as the basis for identifying those aspects for which further research appears to be most valuable. This was considered and is presented as part of Chapter 10.

Chapter 10

Identification of research priorities:
value of information analysis

Decisions about whether to adopt a specific PND identification strategy based upon existing information is uncertain, and there will be a chance that the wrong decision will be made regarding its adoption in the NHS. If the wrong decision is made, then there will be a cost in terms of the health benefit and resources forgone. In this case, women with PND will continue to be screened, and resources will be expended on ineffective identification strategies. Given the prevalence of PND and the major adverse consequence in terms of lost productivity and excess health care utilisation,[242] the consequences of a wrong decision are likely to be substantial. This chapter considered the implications of the uncertainty associated with the cost-effectiveness of identification strategies for women with PND by undertaking value-of-information analysis (VOI). This analysis produced an upper limit to the value of future research that could be undertaken to reduce the uncertainty associated with a decision related to the use of identification strategies for PND in the NHS. VOI analysis provided a formal quantitative approach to establishing whether further primary research appeared to be warranted and also provided an approach to targeting where research would be most worthwhile. The results of the VOI analysis were therefore used to assist in prioritising future research in relation to this decision and to identify particular areas where this appears most valuable[231].

Better quality information produced by further research can help reduce the uncertainty and reduce the chance of a wrong decision being made. Further research, under this analytic perspective, has a value and a benefit to society which can be quantified. The expected costs of decision uncertainty can be interpreted as the Expected Value of Perfect Information (EVPI),[43] since perfect information would eliminate the possibility of making the wrong decision. Furthermore, the EVPI also represents the maximum amount that a decision-maker should be willing to pay for additional evidence to inform this decision in the future. In the UK, this EVPI can be expressed for the total population of people with PND who stand to gain from improved recognition

and management. EVPI was used to provide an upper bound on the value of additional research to that provided by the model. This valuation was then used as a necessary requirement for determining the potential efficiency of further primary research. Applying this decision rule, if the value of this information exceeds the expected costs of conducting further research, then it very quickly becomes apparent that this research is cost effective and a sensible use of finite research resources [43,231]. Conversely, if the value of this information is less than the expected costs of conducting further research, then this research would not be considered cost effective. In addition to providing a global estimate of the total cost of uncertainty related to all inputs in the model, EVPI was also estimated for individual parameters (and groups of parameters) contained in the model. The objective of this analysis (termed partial EVPI) was to identify the model parameters where it would be most worthwhile obtaining more precise estimates. For example, it may be thought that a priority for further research into PND would be to develop a new instrument to replace the EPDS. However, this might not be the most efficient and sensible use of finite NHS research funds. It might be that a randomised controlled trial of an existing instrument, coupled with organisational enhancements of clear decision support systems might yield more informative information. Another alternative is that observational data relating to the longer term consequence of unidentified PND might be needed. The EVPI will be used to provide an upper bound on the value of additional research to that provided by the decision model presented in Chapter 9. This valuation can be then be used as a necessary hurdle for determining the potential efficiency of further primary research.

Methods

The expected costs of decision uncertainty could also be interpreted as the EVPI, as perfect information would eliminate the possibility of making the wrong decision. Furthermore, the EVPI also represents the maximum amount that

a decision-maker should be willing to pay for additional evidence to inform this decision in the future. The EVPI was used to provide an upper bound on the value of additional research to that provided by the model. This valuation was then used as a necessary requirement for determining the potential efficiency of further primary research. Applying this decision rule, additional research was only considered if the EVPI exceeded the expected cost of the research. In addition to providing a global estimate of the total cost of uncertainty related to all inputs in the model, the EVPI was also estimated for individual parameters (and groups of parameters) contained in the model. The objective of this analysis (termed partial EVPI or EVPPI) was to identify the model parameters for which it would be most worthwhile obtaining more precise estimates.

The use of Monte Carlo simulation allowed the expected costs of uncertainty associated with the initial adoption decision to be expressed as the proportion of iterations that results in an adoption decision other than that arising from maximising expected cost-effectiveness. The benefits forgone were simply the difference in the costs and outcomes (net benefit) between the optimal strategy for a given iteration and the strategy that was identified as optimal in the adoption decision (i.e. based on the expected cost-effectiveness estimates). The expectation of benefits forgone over all iterations represented the EVPI per individual.

As information can be of value to more than one individual, the EVPI was also expressed for the total population who stand to benefit over the expected lifetime of the programme/technology. If the EVPI for the population of current and future patients exceeded the expected costs of additional research then it was considered potentially cost-effective to conduct further research. The overall VOI for a population was determined by applying the individual EVPI estimate to the number of people who would be affected by the information over the anticipated lifetime of the technology:

$$ EVPI * \sum_{t=1}^{T} \frac{I_t}{(1+r)^t} $$

where I = incidence in period, t = period, T = total number of periods for which information from research would be useful and r = discount rate.

Results

Base case
Population EVPI

Table 41 reports the population EVPI at the separate willingness to pay thresholds (£20,000, £30,000 and £40,000 per QALY) for the different population sizes (representing time horizons of 10 years, 15 years and 20 years) considered. The EVPI estimates ranged between £5.79 million and £170.05 million across the separate scenarios. Assuming a 10-year horizon, the corresponding estimates were in the region of £5.79 million to £99.51 million, demonstrating that there appears to be considerable value surrounding further research which could reduce the current decision uncertainty represented by the model.

Each population was calculated using the formula given previously, with I_t representing the number of new mothers per annum and assuming a discount rate, r, of 3.5%.[235] An estimate of the number of new mothers per annum was calculated using data from the Office for National Statistics[243] by taking the total number of live births in 2006 (669,601) and subtracting the total number of multiple births in 2006 (10,137). The discounted 10-year, 15-year and 20-year populations were estimated to be 5,676,459, 7,861,154 and 9,700,608 respectively.

As *Figure 41* shows, the population EVPI was negligible for WTP thresholds below approximately £10,000; at these thresholds the strategy of routine case detection was almost certain to be cost-effective and so there appeared to be very little value in obtaining further information about each parameter evaluated as part of the decision model. For WTP thresholds from approximately £10,000 to £41,103 the EVPI for each population increased seemingly exponentially as it became more likely that a strategy other than routine case detection was cost-effective. The growth rate of each population EVPI reached a local maximum at a WTP threshold of £41,103, where the cost-effective strategy switched to the EPDS with a cut point of 16. The growth rate of each population EVPI then slowly increased with WTP threshold and reached a second local maximum at a WTP threshold of around £72,000. At higher WTP thresholds each population EVPI continued to increase, reflecting the greater value placed on health outcomes and hence the greater value placed on perfect parameter information, but the growth rate of each remained relatively flat – at these thresholds it was highly probable that the cost-effective strategy was an identification strategy with a lower cut point (12
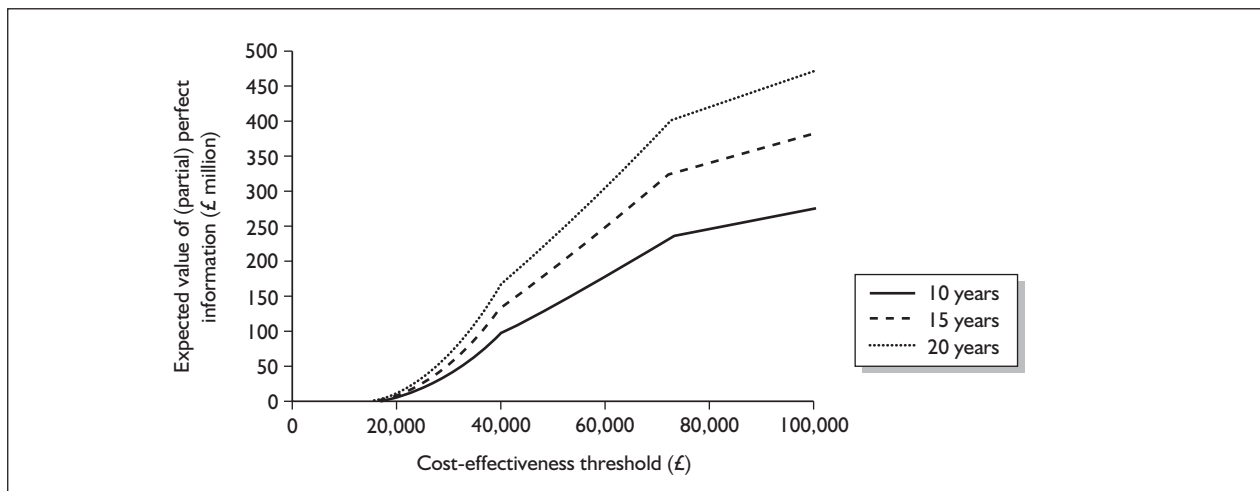
FIGURE 41 Base-case population EVPI.

or below), all of which had very similar expected QALY outcomes.

Table 40 reports the population EVPI at three separate thresholds for the different population sizes considered. The EVPI estimates ranged between £5.79 million and £170.05 million across the separate scenarios. Assuming a 10-year horizon the corresponding estimates ranged from £5.79 million to £99.51 million demonstrating that there appears to be considerable value surrounding further research, which could reduce the current decision uncertainty represented by the model.

Partial EVPI

Although the estimates of the total population EVPI provided a useful global measure of the uncertainty surrounding the choice of identification strategy and the maximum value of future research aimed at reducing such uncertainty, they did not provide any indication of which particular aspects future research would be of most value targeting. Indeed, the population EVPI was only relevant in terms of informing further research that could address each of the separate aspects (e.g. the separate elements associated with the diagnostic and treatment models) simultaneously (e.g. by undertaking a prospective study addressing

both the effectiveness and cost-effectiveness of identification and treatment strategies). However, the EVPPI could be used to estimate the maximum value of reducing uncertainty around particular parameters (or groups of policy-related parameters), allowing future research to be more specifically targeted at those parameters for which more precise estimates would be most valuable and for which a range of different study designs and approaches may be possible.

Table 41 presents the EVPPI for a number of groups of parameters which were considered to represent groupings that were relevant to both broader policy questions and the actual design of further research. For example, the diagnostic test performance of the identification strategies (as a whole and then separately according to the instrument) was separated from other aspects related to the diagnostic model (i.e. epidemiological parameters such as the prevalence of depression and the probability of routine case detection) and to the subsequent treatment model and the estimates of quality of life assigned in the model. This separation was undertaken both to reflect the different uncertainties that exist in these separate elements (and their individual contributions to the overall costs of current decision uncertainty)

TABLE 40 *Base-case population EVPI*

Population	EVPI for maximum WTP		
	£20,000	**£30,000**	**£40,000**
10 years	£5,789,989	£40,075,803	£99,508,333
15 years	£8,018,377	£55,499,747	£137,806,029
20 years	£9,894,620	£68,486,294	£170,051,662
EVPI, expected value of perfect information; WTP, willingness to pay.			

TABLE 41 *Base-case EVPPI*

| Parameter group | EVPI/EVPPI for max. WTP (10-year population) | | |
	£20,000	£30,000	£40,000
Total population EVPI (all parameters)	£5,789,989	£40,075,803	£99,508,333
EPDS and BDI (all cut points) sensitivity/specificity	£0	£3,178,817	£55,742,831
EPDS (all cut points) sensitivity/specificity	£0	£2,894,994	£52,393,720
BDI (all cut points) sensitivity/specificity	£0	£227,058	£12,147,623
Utility weights	£0	£170,294	£7,209,103
Other diagnostic parameters	£0	£0	£5,279,107
Treatment parameters	£227,058	£9,025,570	£40,075,803

BDI, Beck Depression Inventory; EPDS, Edinburgh Postnatal Depression Scale; EVPI, expected value of perfect information; EVPPI, expected value of partial perfect information; WTP, willingness to pay.

and because these separate aspects suggest different research studies and designs (i.e. the study designs and associated costs are likely to be markedly different for the parameters of the treatment model, which would ideally be informed by a randomised trial to minimise potential bias, compared with those represented by the utility parameters, for which issues of bias may be less important in terms of designing further research, i.e. an observational study may be considered appropriate). The estimates of the EVPPI for these groups of parameters were reported for three WTP thresholds assuming a 10-year population. For comparative purposes the total population EVPI is also given.

For a WTP threshold of £20,000 per QALY, over a 10-year population, the EVPPI around the parameters associated with the treatment model alone was £227,058. There appeared no economic value in obtaining additional information solely around any of the other groups of parameters (although in relative terms there appeared significant value in obtaining perfect information around all parameters simultaneously). This finding was not unexpected as in the base-case model the probability that formal identification strategies were cost-effective was very low (and hence there was low uncertainty that routine care was optimal). However, the ICER associated with the treatment option considered (£17,481 per QALY for structured psychological therapy compared with usual care) was close to the £20,000 threshold and hence there appeared to be a high cost of decision uncertainty in relation to the question of whether treatment was itself cost-effective regardless of the identification approach employed.

For higher WTP thresholds (£30,000–40,000) there was positive expected value associated with obtaining perfect information around each of the groups of parameters separately, most notably the sensitivity and specificity associated with the diagnostic performance of the alternative identification approaches (from £3.18 million to £55.74 million) and the treatment model parameters (from £9.03 million to £40.08 million). In terms of the diagnostic performance parameters, the value associated with the EPDS (from £2.89 million to £52.39 million) represented the majority of the value associated with the alternative identification approaches considered in the model. This was because one of the key determinants of a parameter's EVPPI was the likelihood that more precise estimates of its true value would result in a change in the cost-effective strategy (this was also true for groups of parameters).

Sensitivity analysis: scenario 3
Partial EVPI results including Whooley questions

Absent from the previous discussion of the EVPPI was the value of perfect information around the sensitivity and specificity of the Whooley questions. This was because the base case did not include this identification strategy. Given the potential policy importance of this strategy in light of the recent NICE guidance, the EVPPI estimates were re-estimated from the scenario that included the Whooley questions as an alternative identification strategy. The population EVPI and EVPPI for each group of parameters including the Whooley questions are reported in *Table 42*. Given the potential concerns noted elsewhere as to the

TABLE 42 *EVPPI when the Whooley questions are considered as an alternative identification strategy*

Parameter group	EVPI/EVPPI for max. WTP (10-year population)		
	£20,000	£30,000	£40,000
Total population EVPI (all parameters)	£5,562,930	£39,678,451	£101,438,325
All screens sensitivity/specificity	£0	£3,178,817	£56,253,712
EPDS (all cut points) sensitivity/specificity	£0	£3,008,523	£52,166,662
BDI (all cut points) sensitivity/specificity	£0	£227,058	£11,750,271
Whooley questions sensitivity/specificity	£0	£0	£1,135,292
Utility weights	£0	£113,529	£13,055,857
Other diagnostic parameters	£0	£56,765	£11,409,683
Treatment parameters	£340,588	£14,985,853	£64,711,637

BDI, Beck Depression Inventory; EPDS, Edinburgh Postnatal Depression Scale; EVPI, expected value of perfect information; EVPPI, expected value of partial perfect information; WTP, willingness to pay.

validity of the diagnostic data in the context of women with PND, both the results of the cost-effectiveness analysis and the VOI results for this should be treated with some caution.

Including the Whooley questions did not radically alter the magnitude of the population EVPI or the ranking in terms of the relative value of the different groups of parameters considered. Interestingly, the EVPPI for the sensitivity and specificity of the Whooley questions was zero for WTP thresholds of £20,000 and £30,000 and relatively small for a WTP threshold of £40,000. The EVPPI for the EPDS and BDI identification strategies remained approximately the same as those in the base case, whereas the EVPPI for the utility weights, diagnostic parameters and treatment parameters were similar to those of the base case for WTP thresholds of £20,000 and £30,000 but were much greater for a WTP threshold of £40,000.

Discussion

The results from Chapter 9 indicated that there was a considerable decision uncertainty surrounding the role of formal identification strategies for the management of women with PND. This finding was reflected in the high cost of decision uncertainty represented by the EVPI estimates presented here suggesting that further research is potentially worthwhile. At low thresholds of cost-effectiveness, future research appeared most worthwhile targeted at evidence related to the effectiveness and cost-effectiveness of treatment strategies for the management of women with confirmed PND. At higher thresholds of cost-effectiveness

there appeared markedly higher potential value associated with further research more generally and also specifically around: (1) diagnostic test performance (primarily related to the use of the EPDS); (2) treatment strategies for confirmed PND; (3) the impact of PND on HRQoL; and (4) other epidemiological data considered in the diagnostic model (e.g. prevalence rates, routine case detection).

Although these findings indicate that further research is potentially worthwhile, several factors need to be taken into consideration. First, these values represent an upper bound to the value of further research as a whole and in relation to the individual parameter groups as they represent the value of perfect information (i.e. assuming that further research will completely resolve any remaining uncertainties). Clearly further research would only partially resolve these uncertainties; the costs reported here are thus only indicative as to the potential value of further research. In effect, they represent a necessary, but not sufficient condition for further research to be considered efficient (assuming that the costs of research were actually lower than that represented by the EVPI estimates). Second, these estimates are based on the assumptions and strategies considered in the associated decision model. The model itself focused on strategies for which there were considered sufficiently robust data for them to be included in the evaluation. As such the full range of potentially feasible strategies was not considered (e.g. although a number of separate cut points was considered for the EPDS, only a single cut point was considered for the BDI). Hence, the finding that the majority of the value associated with the diagnostic performance of the alternative

identification strategies was attributed to the use of the EPDS (and not the BDI/Whooley questions) needs to be weighed against the more restrictive range of cut points considered in the case of the BDI and the validity of the diagnostic performance data based on the Whooley questions. Hence, there may still be considerable value that could be associated with obtaining more reliable data from a range of alternative diagnostic approaches not considered as part of the decision modelling work.

Chapter 11

Should identification of postnatal depression be implemented as a national screening policy according to National Screening Committee criteria?

The NSC advises ministers and the NHS in the UK about all aspects of screening policy. In 1996 the NHS was instructed not to introduce any new screening programmes until the NSC had reviewed their effectiveness. The NSC uses research evidence, pilot programmes and economic evaluations to assess programmes against a set of internationally recognised criteria (*Table 43*). Screening programmes are assessed to ensure that the screening does more good than harm at a reasonable cost.

Screening for PND was evaluated against the NSC criteria by Dr Judy Shakespeare in 2001[26] and by the Child Health Subgroup in 2002. The EPDS was identified as the most frequently advocated identification strategy and many of the items were assessed in relation to this tool. It was recommended that, until more research is conducted into its potential for routine use in screening for PND, the EPDS should not be used as a screening tool.

The previous reviews of PND screening against NSC criteria have stimulated substantial debate and have left practitioners and policy-makers in a state of uncertainty regarding whether or not to screen in this population. The programme of research presented here allows, within an explicit evidence-based framework, many of the areas of uncertainty highlighted in the previous NSC report to be explored. Hence, we revisit and revise the examination of PND screening against NSC criteria in the light of the results of our evidence synthesis and decision modelling. The following key areas were assessed:

- the nature of the test (with particular reference to acceptability and validity)
- the availability of effective treatments
- the effectiveness of the screening programme (with reference to clinical and cost-effectiveness).

Items 5, 6, 7, 13 and 16 were revisited in light of the results from the evidence synthesis and decision model presented in previous chapters.

Item 5: There should be a simple, safe, precise and validated screening test

Numerous potentially suitable identification strategies for PND were identified. The systematic review in Chapter 5 identified that the EPDS was the most frequently used instrument and was the only instrument for which sufficient data were available to combine studies at multiple cut points. In terms of test performance, the EPDS performed reasonably well: sensitivity ranged from 0.60 (specificity 0.97) to 0.96 (specificity 0.45) for major depression only; from 0.31 (specificity 0.99) to 0.91 (specificity 0.67) for major or minor depression; and from 0.38 (specificity 0.99) to 0.86 (specificity 0.87) for any psychiatric disorder. Although the EPDS has reasonable sensitivity and specificity, some women with PND will be unidentified and some women without PND will be wrongly identified as having PND.

Criterion met? Yes, for the EPDS. There was a lack of evidence for the other potential identification strategies identified.

Item 6: The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed

A wide variety of cut points for the EPDS have been reported. There were sufficient data at 10 cut points to be able to combine studies for certain types of disorders. In the original validation

TABLE 43 *National Screening Committee criteria*

The condition

1   The condition should be an important health problem

2   The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor, disease marker, latent period or early symptomatic stage

3   All the cost-effective primary prevention interventions should have been implemented as far as practicable

4   If the carriers of a mutation are identified as a result of screening the natural history of people with this status should be understood, including the psychological implications

The test

5   There should be a simple, safe, precise and validated screening test

6   The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed

7   The test should be acceptable to the population

8   There should be an agreed policy on the further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals

9   If the test is for mutations the criteria used to select the subset of mutations to be covered by screening, if all possible mutations are not being tested, should be clearly set out

The treatment

10   There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment

11   There should be agreed evidence-based policies covering which individuals should be offered treatment and the appropriate treatment to be offered

12   Clinical management of the condition and patient outcomes should be optimised in all health-care providers prior to participation in a screening programme

The screening programme

13   There should be evidence from high-quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity. Where screening is aimed solely at providing information to allow the person being screened to make an 'informed choice' (e.g. Down syndrome, cystic fibrosis carrier screening), there must be evidence from high-quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened

14   There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/intervention) is clinically, socially and ethically acceptable to health professionals and the public

15   The benefit from the screening programme should outweigh the physical and psychological harm (caused by the test, diagnostic procedures and treatment)

16   The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (i.e. value for money)

17   There should be a plan for managing and monitoring the screening programme and an agreed set of quality assurance standards

18   Adequate staffing and facilities for testing, diagnosis, treatment and programme management should be available prior to the commencement of the screening programme

19   All other options for managing the condition should have been considered (e.g. improving treatment, providing other services), to ensure that no more cost-effective intervention could be introduced or current interventions increased within the resources available

20   Evidence-based information, explaining the consequences of testing, investigation and treatment, should be made available to potential participants to assist them in making an informed choice

21   Public pressure for widening the eligibility criteria, for reducing the screening interval and for increasing the sensitivity of the testing process should be anticipated. Decisions about these parameters should be scientifically justifiable to the public

22   If screening is for a mutation the programme should be acceptable to people identified as carriers and to other family members

study of the EPDS[16] a cut point of 10 for 'possible depression' and a cut point of 13 for 'probable depression' were suggested. The lower the cut point used to distinguish between cases and non-cases the higher the sensitivity becomes. Increasing the sensitivity of an identification strategy will lead to fewer women with PND being unidentified. Unfortunately, increasing sensitivity results in lower specificity values leading to an increase in the number of women wrongly diagnosed with PND and increasing demands on NHS resources. The systematic review in Chapter 5 demonstrated that the optimal cut point, in terms of the trade-off between sensitivity and specificity, was 12 for major depression only, 10 for major or minor depression and 9 for any psychiatric disorder. If the cut point was chosen to maximise sensitivity (from a clinical perspective) then from this analysis the optimal cut point was 7 for major depression only, 8 for major or minor depression and 9 for any psychiatric disorder. From an economic perspective the results suggested that in the scenarios considered the most cost-effective identification approach would be the EPDS at a cut point of 10 or higher.

Criterion met? Yes, in principle for the EPDS. There was a lack of evidence for other potential identification strategies identified.

Item 7: The test should be acceptable to the population

The systematic review in Chapter 6 identified 16 studies that explored the acceptability of methods to identify PND. The most frequently explored views of women and health professionals were those regarding the EPDS. Overall, the majority of studies indicated that the EPDS was acceptable when undertaken in the home, with due attention to training, with empathetic skills of the health visitor and with due consideration of positive responses to question 10 ('the thought of harming myself has occurred to me').

Criterion met? Yes for the EPDS. There was a lack of evidence for other potential identification strategies identified.

Item 13: There should be evidence from high-quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity

The systematic review in Chapter 7 highlighted that insufficient evidence is available to conclude that identification strategies are effective in improving maternal and infant outcomes. Some suggestive evidence indicated that the EPDS, maybe with some enhancement of care, may lead to reductions in the number of women with EPDS scores above a certain threshold or reductions in EPDS scores. Despite additional outcomes being considered, only EPDS outcomes were presented across all of the studies included in the review.

Criterion met? No.

Item 16: The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (i.e. value for money)

No full economic evaluations of PND identification strategies were identified in the systematic review in Chapter 8. In the absence of existing cost-effectiveness studies of PND identification strategies, a decision-analytic model was developed. The results of the base-case analysis suggested that the use of formal identification strategies do not appear to represent value for money based on conventional thresholds of cost-effectiveness used in the NHS. However, the scenarios considered

121

demonstrated that this conclusion was primarily driven by the costs of false positives assumed in the base-case model. Alternative assumptions employed in separate scenarios result in more favourable estimates of cost-effectiveness, such that the use of the EPDS as an identification strategy to identify women with PND falls within these conventional thresholds. For example, when the cost of a false-positive diagnosis was assumed to be the cost of a single GP attendance, the EPDS using a cut point of 10 emerged as the optimal strategy in terms of cost-effectiveness. A definitive answer to the question as to whether formal identification strategies are cost-effective, and, if they are, which individual strategy is optimal in cost-effectiveness terms, clearly requires further more reliable evidence.

Criterion met? No.

Reflection on current policy and practice within the UK

We found that the accepted criteria for a PND screening programme were not currently met by any of the identification strategies identified. The evidence suggested that the EPDS is a simple, safe, precise and validated screening test, that in principle a suitable cut-off level could be defined and that the test is acceptable to the population. Evidence surrounding the clinical and cost-effectiveness of PND screening with the EPDS is lacking. There was insufficient evidence for all other identification strategies identified to assess them against the NSC criteria. As outlined in Chapter 1 current NICE guidance recommends the use of two questions to identify possible depression and a third question if the women answers 'yes' to either of the initial questions [(1) 'During the past month, have you often been bothered by feeling down, depressed or hopeless?', (2) 'During the past month, have you often been bothered by little interest or pleasure in doing things?' and (3) 'Is this something you feel you need or want help with?']. It is worth noting that no evidence was identified across the four systematic reviews, in terms of validity, acceptability, clinical effectiveness and cost-effectiveness, for these three questions in a postnatal population. Thus, we would also conclude that the NSC criteria for a PND screening programme using the three questions would not currently be met.

Chapter 12
Discussion and conclusions

This was a substantial review that required the application of innovative review methods in a new and exciting area – particularly reviews of psychometric properties, integrating qualitative and quantitative research findings, the effectiveness of case finding/diagnostic interventions and the use of decision modelling and VOI analysis. The findings from this body of research are discussed below in terms of how each of the objectives outlined in Chapter 2 has been addressed and the conclusions that can be drawn.

Objective 1: To provide an overview of all available methods to identify postnatal depression in primary care and to assess their validity

The survey of methods to identify PND highlighted that there were numerous measures that could be used. Diagnostic measures such as the DSM, ICD and RDC were considered the gold standard reference case for the classification of mood disorders within the postnatal period. However, the current classification of PND as a mood disorder within the postnatal period reflects underlying uncertainty regarding the entity of PND as a distinct diagnosis. In addition, disparities in diagnosis may have arisen because of differences in the time frame specified for onset and differences in the criteria for a major depressive disorder between the classification systems, for example more symptoms must be established using the RDC than using the DSM. In contrast to the use of diagnostic interview schedules, four other approaches to identify PND were identified: clinician-rated scales, generic depression identification strategies, specific PND identification strategies and case-finding questions. Clinician-rated scales were defined as measures of depression used to standardise clinical judgements and provide ratings of the duration and severity of symptoms. Generic depression (and sometimes anxiety) instruments were those designed and validated for the identification of depression in non-postnatal populations, whereas postnatal-specific measures were those designed

and validated for the identification of depression in postnatal populations. Both generic depression measures and postnatal-specific measures assessed self-reported depressive symptoms and subjects rated their symptoms in terms of their frequency and severity. Six postnatal-specific measures were identified. Two of these measures (PRQ and PI) were developed for use prenatally to identify those women with depression during pregnancy and to identify those women at risk of development of significant depression in the postnatal period. In some areas of research there has been shift away from using self-report strategies to using case-finding questions (e.g. for depression). Recent NICE guidance issued on antenatal and postnatal mental health recommended using case-finding questions developed by Whooley and colleagues with an additional help question if women responded 'yes' to one of the two questions.

All potential PND identification strategies identified from the survey were then subjected to a systematic review of their validity (in terms of key psychometric properties). In total, 14 identification strategies were validated among women during pregnancy or the postnatal period (up to 1 year). Identification strategies included PND-specific identification strategies (EPDS, PDSS, PRQ, PI), generic depression identification strategies (BDI, GHQ, HADS, HSCL, HAMD, Zung's SDS, SCL-90-R, Raskin, MADRS) and others (EPDS–GHQ double test). No validation studies were identified that validated the case-finding questions recommended in recent NICE guidance in a postnatal population. By far the most frequently used identification strategy was the EPDS. Studies that reported the results of applying the same identification strategy using the same cut point to diagnose the same type of disorder were pooled using a bivariate meta-analysis. There were sufficient data from postnatal studies across a large number of cut points of the EPDS to be able to combine results and produce pooled summary estimates of sensitivity and specificity. However, there were insufficient data at each cut point for most other identification strategies to be able to pool data. For major depression only the pooled estimates of sensitivity and specificity for the EPDS ranged from 0.60 to 0.96 and from 0.45

to 0.97 respectively; for any psychiatric disorder the pooled estimates of sensitivity and specificity for the EPDS ranged from 0.38 to 0.86 and from 0.85 to 0.99 respectively; and for major or minor depression the pooled estimates of sensitivity and specificity for the EPDS ranged from 0.31 to 0.91 and from 0.67 to 0.99 respectively. In addition, for major or minor depression there were sufficient data to pool the BDI and HAMD results at a single cut point. Results from this analysis highlighted that generic identification strategies may be less sensitive than the EPDS but more specific. When the psychometric attributes were pooled across the studies, high levels of between-study heterogeneity were identified in most analyses (major depression: 0–93%; major or minor depression: 41–98%). Unfortunately, none of the a priori sources of heterogeneity was predictive in a meta-regression analysis, and high levels of between-study heterogeneity remained in the model. There was some suggestive evidence that the timing of administration of the EPDS (within 6 weeks postnatally or not) may be an important factor in influencing diagnostic performance. Two other variables, verification bias and blinding, also demonstrated some potential effects on diagnostic performance. The identification strategies reviewed here appear to be able to identify PND in women during pregnancy and the postnatal period with a degree of accuracy that is similar, if not slightly better, to that for identifying depression in general populations. In an evaluation of case-finding instruments for identifying patients with major depression or dysthymia in primary care, 16 instruments were assessed in 38 studies and the overall sensitivity was 0.79 (95% CI 0.74 to 0.83) and overall specificity was 0.75 (95% CI 0.70 to 0.81). Equivalent estimates from this review for the EPDS resulted in an overall sensitivity of 0.86 (95% CI 0.81 to 0.89) and overall specificity of 0.87 (95% CI 0.80 to 0.92). In summary, the EPDS is the most frequently reported identification strategy and its diagnostic performance seems reasonably good.

Objective 2: To assess the acceptability of methods to identify postnatal depression in primary care

Both qualitative and quantitative methods were used to address the overall question of whether postnatal identification strategies were acceptable to women and health professionals. A textual narrative approach was used to synthesis the evidence, which involved grouping studies together into subgroups, writing a commentary on key aspects of studies in relation to the subgroup and then conducting subgroup synthesis. Most qualitative studies highlighted that English-speaking women thought that the EPDS was acceptable, although one study that used in-depth interviews found that 21 of 39 English-speaking women felt that the EPDS was unacceptable. With regards to the quantitative research, two surveys, comprising several hundred Australian women, reported that the majority of women found that using the EPDS was 'comfortable' to 'very comfortable' or that it was 'fairly easy to complete'.

An important theme about the acceptability of postnatal identification strategies to women and health professionals was to ensure that the women felt comfortable and relaxed about the process so that subsequently they would answer the questions honestly. To achieve this there was evidence to support informing women well in advance of administering the identification strategy, at antenatal visits, that this was going to happen. The evidence also supported the view that identification strategies should be administered in women's homes where there would be more privacy to discuss their emotional well-being. When administering the questionnaire it was suggested that this should be undertaken at around 6–8 weeks postnatally to allow women a period of adjustment after becoming a mother. Evidence also highlighted that rather than the completion of a questionnaire being only a pen and paper assessment it should be an opportunity to open up a dialogue and discuss the results with women as an adjunct to clinical practice. It was important that the health professional was caring and showed an interest and that it was clear to the woman that they were there to support her and not as a threat to taking the baby away. Training in counselling may therefore be beneficial so that the health professional has the skills to deal with the disclosures that the work uncovers. Similarly, health professionals should be sensitive to different cultural attitudes towards being a mother, and the ambiguity of the question about self-harm in the EPDS should be altered. An identification strategy that applies these elements to the routine assessment of women postnatally should make the experience of identification more 'normal' and less anxiety-provoking so that women are more likely to be honest about how they truly feel and therefore more likely to receive the support that they need.

Objective 3: To assess the clinical effectiveness of methods to identify postnatal depression in improving maternal and infant outcomes in primary care

To meet objective 3 we reviewed studies that focused on whether the routine use of case identification strategies for PND or the integration of case identification strategies with enhancements of care resulted in improvements in maternal and infant outcomes. Five studies were identified that compared use of an identification strategy with or without enhancement of care or feedback of scores with either not using an identification strategy or usual care (i.e. providing level I or level II evidence). As identified in the other reviews presented in this report, the EPDS was the most frequently used identification strategy. All of the studies indicated beneficial effects of using the EPDS in reducing EPDS scores, although some of the individual studies did not show statistically significant differences. Studies reporting dichotomous outcomes were combined in a fixed-effects meta-analysis and the pooled estimate gave an OR of 0.64 (95% CI 0.52 to 0.78). Thus, the odds of scoring above the threshold for depression in a population in which a formal method was used to identify PND (with an intervention for those identified) was 0.64 times the odds of scoring above the threshold for depression in a population in which there was no formal method to identify PND (with an intervention for those identified). In total, 25 studies were identified that reported using identification strategies at the recruitment stage to identify women at risk of PND with the aim of examining the impact of various interventions compared with usual care (i.e. level III evidence). Despite a large number of studies being identified there were a number of clinical and methodological differences between the studies, which did not permit statistical pooling of results. Furthermore, it was hard to distinguish between the benefits of using identification strategies and the effects of the intervention under study. Thus, it was difficult to draw conclusions regarding the impact of using identification strategies on maternal and infant outcomes.

Objective 4: To assess the cost-effectiveness of methods to identify postnatal depression in improving maternal and infant outcomes in primary care

Despite an extensive systematic search of the literature none of the studies identified presented full economic evaluations of PND identification strategies. In the absence of existing cost-effectiveness studies of PND identification strategies, a decision-analytic model was developed to evaluate the cost-effectiveness of a range of alternative identification strategies. The model provided a framework for the synthesis of diagnostic accuracy data reported in the validation review with a range of other relevant parameters required to establish the cost-effectiveness of using formal identification strategies for women with PND. In the base-case analysis it was assumed that a single identification strategy was used at 6 weeks postnatally to identify women with major or minor depression. The choice of identification strategies was limited to the EPDS (cut points 7–16) and the BDI (cut point 10). Separate scenarios were considered subsequent to the base-case analysis including alternative classifications (i.e. considering major depression only) and alternative identification strategies (considering a separate scenario including the Whooley questions and a scenario in which a separate confirmatory strategy was used, employing a gold-standard instrument such as SCID, in women identified as positive cases from the results of the initial diagnostic instrument). The results of the base-case analysis suggested that the use of formal identification strategies did not appear to represent value for money based on conventional thresholds of cost-effectiveness used in the NHS. However, the scenarios considered demonstrated that this conclusion was primarily driven by the costs of false positives assumed in the base-case model. Alternative assumptions employed in separate scenarios resulted in more favourable estimates of cost-effectiveness, such that the use of the EPDS as an identification strategy to identify women with PND considered in some of these scenarios fell within these conventional thresholds. For example, when the cost of a false-positive diagnosis was assumed to be that of a single GP attendance, the EPDS using a cut point of 10 emerged as the optimal strategy in terms of cost-effectiveness. Interestingly, this corresponded closely with the

125

results presented in the validation review in which the trade-off between sensitivity and specificity was considered (but not in terms of their related cost and outcomes). A definitive answer to the question of whether formal identification strategies are cost-effective, and, if they are, which individual strategy is optimal in cost-effectiveness terms, clearly requires further more reliable evidence in relation to the costs of managing false positives. However, the results presented here suggest that, in the scenarios considered, the most cost-effective identification approach would be the EPDS at a cut point of 10 or higher.

Objective 5: To identify research priorities and the value of further research into methods to identify postnatal depression from the perspective of the UK NHS

The decision-analytic model developed to meet the previous objective was also used to provide a vehicle for identifying potential future research priorities by undertaking a VOI analysis. The VOI analysis produced an upper limit to the value of future research that could be undertaken to reduce the uncertainty associated with a decision related to the use of identification strategies for PND in the NHS. The results from the decision model indicated that there was a considerable decision uncertainty surrounding the role of formal identification strategies for the management of women with PND. This finding was reflected in the high cost of decision uncertainty represented by the EVPI estimates suggesting that further research is potentially worthwhile. At low thresholds of cost-effectiveness, future research appeared most worthwhile targeted at evidence relating to the effectiveness and cost-effectiveness of treatment strategies for the management of women with confirmed PND. At higher thresholds of cost-effectiveness there appeared markedly higher potential value associated with further research more generally and also specifically around (1) diagnostic test performance (primarily related to the use of the EPDS); (2) treatment strategies for confirmed PND; (3) the impact of PND on HRQoL; and (4) other epidemiological data considered in the diagnostic model (e.g. prevalence rates, routine case detection).

Although these findings indicated that further research was potentially worthwhile, several factors need to be taken into consideration. First, these

values represent an upper bound to the value of further research as a whole and in relation to the individual parameter groups as they represent the value of perfect information (i.e. assuming that further research will completely resolve any remaining uncertainties). Second, these estimates were based on the assumptions and strategies considered in the associated decision model. The model itself focused on strategies for which there was considered sufficiently robust data for them to be included in the evaluation. As such, the full range of potentially feasible strategies was not considered (e.g. although a number of separate cut points was considered for the EPDS, only a single cut point was considered for the BDI). Hence, the finding that the majority of the value associated with the diagnostic performance of the alternative identification strategies was attributed to the use of the EPDS (and not the BDI/Whooley questions) needs to be weighed against the more restrictive range of cut points considered in the case of the BDI and the validity of the diagnostic performance data based on the Whooley questions. Hence, there may still be considerable value that could be associated with obtaining more reliable data from a range of alternative diagnostic approaches not considered as part of the decision modelling work.

Objective 6: To assess whether methods to identify postnatal depression meet minimum criteria outlined by the NSC in the light of the evidence synthesis

In the light of the results of our evidence synthesis and decision modelling we revisited the examination of PND screening against five of the NSC criteria. We found that the accepted criteria for a PND screening programme were not currently met. The evidence suggested that there is a simple, safe, precise and validated identification strategy, that in principle a suitable cut-off level could be defined and that the strategy is acceptable to the population. Evidence surrounding the clinical effectiveness and cost-effectiveness of methods to identify PND is lacking.

Limitations, assumptions and uncertainties

A series of systematic reviews was undertaken using innovative methodological approaches to summarise the available evidence. These are the

largest and most comprehensive reviews to have been undertaken within the area of PND to date. Systematic searches were undertaken using 20 electronic databases, forward citation searching of key literature, personal communication with authors and searching of reference lists. All databases were searched from their inception until February 2007 with no language or other restrictions being applied. A large number of potentially relevant articles were identified from the searches ($n = 11,945$), of which 125 were selected for full assessment. Because of the large number of studies identified it is acknowledged that some articles of relevance may have been excluded unintentionally; however, two reviewers independently assessed study titles and abstracts so the likelihood of this was minimised. The comprehensive nature of the review was highlighted by the fact that 64 validation studies, 16 acceptability studies and 30 clinical effectiveness studies were included in reviews 1, 2 and 3 respectively. The number of studies identified, given previous research published in this area, exceeded the authors' expectations. Despite an extensive search of the literature, which included an additional focused search of economic databases, no full economic evaluations of methods to identify PND were identified.

As the searches were undertaken in February 2007 it is possible that new literature will have emerged while preparing the HTA report for final publication. To assess the magnitude of literature published and to assess how any new literature might impact on the results of our review we undertook a scoping search of MEDLINE. The MEDLINE search outlined in Appendix 1 was re-executed in January 2009. After deduplicating and excluding studies included in the previous searches 844 studies were retrieved. After screening titles and abstracts 37 appeared to be potentially eligible for inclusion. On closer inspection 21 of the 37 studies, including three studies that were already included in the review from correspondence with authors or in other formats, would not have met the inclusion criteria. Hence, 16 studies would have been eligible for inclusion in one of the systematic reviews: validation ($n = 11$), acceptability ($n = 2$), clinical effectiveness ($n = 3$) and cost-effectiveness ($n = 0$).

Eleven studies[244–254] assessed the validity of identification strategies, seven of which assessed the validity of the EPDS compared with a diagnostic interview conducted according to internationally recognised criteria. Given the large

number of studies assessing the EPDS already included in the validation review the addition of these new studies is unlikely to alter the conclusions drawn from Chapter 5. The validity of nine other instruments was assessed: PDSS ($n = 2$); Postpartum Depression Risk Scale (PDRS; $n = 1$); BDI ($n = 1$); K6 ($n = 1$); K10 ($n = 1$); PHQ-9 ($n = 1$); Self-Reporting Questionnaire (SRQ; $n = 1$); How I feel ($n = 1$); Aga Khan University Anxiety and Depression Scale (AKUADS; $n = 1$). Although the large majority of the literature focused on the EPDS, it is interesting to note that some new instruments have been validated in women during pregnancy or the postnatal period. Given the small number of studies exploring these new instruments it would not have been possible to combine the sensitivity and specificity data in a meta-analysis.

Two studies[255,256] assessed the acceptability of the EPDS when used as a screening tool and they both found that the EPDS was acceptable. One study focused on the use of the EPDS prenatally and the views of women were explored and the second study focused on the use of the EPDS postnatally and the views of health professionals were sought.

Three studies[257–259] provided level III evidence for the clinical effectiveness review. One of the studies provided additional outcome data for a study already included in the review.[257] The remaining two studies were both treatment studies and used the EPDS during the recruitment stage to identify women at risk of PND with the aim of examining an exercise support programme and a multicomponent intervention compared with usual care. As all of the studies identified provided level III evidence using the EPDS as an identification strategy the conclusions drawn from Chapter 7 are unlikely to change in light of this new literature.

In summary, the scoping search of MEDLINE identified 16 studies that would potentially be eligible for inclusion in one or more of the systematic reviews; however, none of the studies would substantially alter the conclusions drawn from any of the reviews.

Validity of methods to identify postnatal depression: systematic review 1

The QUADAS tool was used to appraise the quality of studies included in the validation review (review 1). There was wide variability in the results across the individual items, although none of the studies fulfilled all of the quality criteria.

Studies demonstrated high quality in five out of the eight questions focusing on bias (questions 3, 5, 6, 7, and 14) and the three questions relating to reporting quality (questions 8, 9 and 13). The poorest quality rating was associated with question 10 regarding whether the index test results were interpreted without knowledge of the reference standard results. This item is important as interpretation of the results of the index test may be influenced by knowledge of the results of the reference standard leading to inflated estimates of diagnostic accuracy. Two other poorly reported items were those relating to selection criteria and disease progression bias. Application of QUADAS caused several difficulties, particularly with reference to the items on uninterpretable results and withdrawals. We encountered problems applying the guidance notes and found that despite the modifications to the guidance it was difficult to apply these items to the studies. It was unclear sometimes whether there were truly any uninterpretable results or withdrawals. Although we followed the guidance for these two items quite prescriptively, after resolving guidance clarity by discussion, we felt uncomfortable that we had to rely on the fact that nothing had been reported in the paper rather than being able to read an explicit statement in the paper that definitively stated that there were no uninterpretable data or withdrawals.

Within review 1 the most sophisticated method of statistically pooling diagnostic accuracy studies, a bivariate meta-analysis, was used to combine studies and produce pooled estimates of sensitivity and specificity. In the majority of validation studies a range of sensitivity and specificity values were recorded for different cut points on the identification strategy used. As multiple data points were presented within each study we pooled data points at individual cut points to overcome this. One of the drawbacks of pooling at individual cut points was that fewer studies were included and this subsequently led to reductions in power. Because of the low number of studies included at some cut points, meta-regression could not be undertaken and potential sources of heterogeneity could not be explored. When there were sufficient numbers of studies to pool psychometric attributes, high levels of between-study heterogeneity were identified in most analyses (major depression: 0–93%, major or minor depression: 41–98%). Unfortunately, none of the a priori sources of heterogeneity was predictive in a meta-regression analysis, and high levels of between-study heterogeneity remained in the model, although the analyses may have lacked the statistical power to detect any systematic differences between the groups.

Acceptability to women and health professionals of methods to identify postnatal depression: systematic review 2

The overall methodological quality of studies included in the acceptability review could not be assessed using a single appraisal tool as evidence was collated from both qualitative and quantitative research. The quality of both types of research was assessed separately. For the quantitative studies, surveys were used and they could be open to selection bias. In particular, only one[179] of the four surveys selected a random sample. The surveys also had poor response rates further indicating the possibility of bias in those women or health professionals who chose to respond and possibly limiting the validity and generalisability of the findings. Most qualitative studies recruited a convenience sample of participants and collected data using semistructured interviews. It is arguable that the sampling strategy did not promote the generalisability of the individuals included in the sample. Moreover, the two studies[167,178] that used in-depth interviews to explore this subject were the most critical about the acceptability of postnatal identification strategies using the EPDS. Most studies collected data around the same time as the process of PND identification, but a criticism of both studies that used in-depth interviews was that the data collection was conducted several months after the process of PND identification. This suggests that there might be recall bias in the response of the participants, or that a particular type of participant was willing to take part in the study when collecting data several months later.

An innovative approach was used to synthesise evidence from qualitative and quantitative studies included in this review. Survey methods were the only quantitative approach used to assess women's or health professionals' views towards the identification strategies, or to record refusal or non-participation rates for the completion of a standardised questionnaire. It was difficult to integrate the evidence from qualitative and quantitative research because the surveys only asked broad questions about, for example, how comfortable women found the process of PND identification. Hence, it was not possible to assess how different themes discussed in the qualitative research were included in the surveys and whether this subsequently affected the estimates of acceptability of identification strategies. Therefore, from the surveys alone it was not possible to understand what made an identification strategy acceptable or not.

A number of strategies have been proposed to improve PND identification, which fall into five broad categories:

- Post-natal identification using specially developed standardised post-natal questionnaires (such as the Edinburgh Post-natal Depression Scale - EPDS16 and the US Postpartum Depression Screening Scale - PDSS).[17]
- Post-natal identification using standardised generic questionnaires for depression (such as the Beck Depression Inventory – BDI).[18]
- Pre-natal screening using standardised depression questionnaires to identify those with pre-existing depression and those at risk of developing significant depression in the post-natal period.[19]
- Pre-natal screening using known risk factors for PND (such as previous history of depression and lack of social support) to identify those who are likely to subsequently develop depression in the post-natal period.[20]
- The use of training packages targeted at healthcare professionals designed to enhance awareness and recognition of clinical signs of post-natal depression and to ensure that a thorough psychosocial assessment is provided on a routine basis.[21]

While we considered all of the PND identification strategies categorised above, none of the studies we identified, and that met our inclusion criteria, focused on pre-natal screening using known risk factors for PND or training packages. The majority of the studies identified evaluated the use of specially developed standardised post-natal questionnaires or standardised depression questionnaires which were administered to women during the post-natal period.

Clinical effectiveness of methods to identify postnatal depression in improving maternal and infant outcomes: systematic review 3

Although 30 studies were included in review 3, only five studies were identified that compared using an identification strategy with or without enhancement of care or feedback of scores with not using an identification strategy or usual care. Across all of the included studies there were a number of methodological weaknesses. We included controlled trials as well as randomised trials and, for example, in some of the studies odd or even expected dates of delivery were used to randomise participants to treatment groups. Obviously, such

methods of randomisation are not truly random and thus, in practice, this often results in selection bias being introduced. In addition, in some of the studies described as RCTs it was often unclear how the randomisation sequence was generated and what methods were used to conceal the sequence. Hence, it was difficult to judge whether the methods used to randomise participants to treatment groups were subject to bias. In some studies randomisation was undertaken on a cluster rather than on an individual basis, but in two out of the three studies the analyses did not account for this. Other frequently occurring problems were associated with attrition, non-compliance or non-attendance and use of intention to treat analysis.

All of the studies used the EPDS as a method to identify PND and the outcome of choice within these studies was depression score, both as a continuous score (i.e. mean score) or as a binary measure (i.e. number of women scoring above and below a threshold for depression). It would have been interesting to explore the impact of using the EPDS on additional maternal and infant outcomes; however, an insufficient number of studies reported such outcomes to undertaken any analyses. Another limitation of the studies included was that measures of dispersion were often not reported, thus analyses were unable to be undertaken with these studies. Finally, it was hard to distinguish between the benefits of using identification strategies, enhancements of care and the effects of any intervention under study. Hence, overall, it was difficult to draw any firm conclusions about clinical effectiveness.

Decision model of methods to identify postnatal depression and value of information

This represents the first attempt to formally evaluate the potential cost-effectiveness of alternative identification strategies for the management of women with PND. Equally important is that this evaluation provides a systematic, integrated and transparent approach to synthesising evidence reported on the diagnostic performance (validation review) with the best available evidence relating to the subsequent management and outcomes reflecting current policy and practice in the NHS. Taken at face value the economic evaluation demonstrated that use of formal identification strategies did not appear to represent value for money based on conventional thresholds of cost-effectiveness used in the NHS. However, the scenarios considered demonstrated

that this conclusion was primarily driven by the costs of false positives assumed in the base-case model.

There was a lack of reliable published data associated with the costs of false positives and hence a conservative approach was employed. It was considered that the additional care costs would involve one community psychiatric nurse visit of 1 hour, three GP visits of 10 minutes each and four health visitor home visits of 45 minutes each. These costs represent a significant additional cost associated with the use of identification strategies that in reality remains highly uncertain. When alternative and plausible estimates (e.g. assuming a single GP appointment) or approaches to managing false positives (e.g. the use of a confirmatory SCID assessment) were considered, markedly more favourable cost-effectiveness estimates for formal identification approaches were demonstrated. Furthermore, the costs of false positives associated with routine care were not considered as part of the base-case analysis and some of the costs attributed to the formal identification strategies would also be incurred during routine practice. Although there was an absence of reliable published data on the costs of false positives, it would appear reasonable to conclude that the actual estimate lies somewhere between the full costs of the additional care considered in the base-case analysis and that assumed for a single GP appointment. Furthermore, there were limited published data available for estimating parameters in the model, namely: the costs of false positives, the probability that PND was identified via routine care at six weeks, the risk of relapse and the utility weights. As a result, the estimates used in the model were derived from studies of general depressed populations (i.e. non postnatal populations), which represents a limitation of the model.

Although a strength of the decision model was that estimates of the diagnostic performance of the identification strategies were taken from the bivariate meta-analysis, undertaken to meet objective 1, it was also a limitation. This is because of the evidence available from the primary studies included in the review. The majority of the research has been focused on the performance of the EPDS. Subsequently, the EPDS was the only identification strategy for which there were sufficient data at more than one cut point to be able to combine results and produce pooled summary estimates of sensitivity and specificity. The impact of this within the decision model was that the performance

of other identification strategies could not be assessed. Finally, it should be noted that the model focused on the costs and outcomes associated with the mother herself. Because of a lack of reliable evidence, no account was taken of the potential impact of successful identification and subsequent management of PND on other family members or the infant. If identification strategies and subsequent management have an important effect on these aspects then the results presented here will represent highly conservative estimates of the potential value of identification approaches. Clearly there remain a number of uncertainties in relation to the assumptions that underpin the model and also in terms of the uncertainty characterised by the probability distributions assigned to the inputs. While the underlying studies provide a partial understanding of the performance of the different tools, there are significant gaps in the evidence that the modelling cannot remove.

Although the findings from the VOI indicated that further research was potentially worthwhile, several factors need to be taken into consideration. First, these values represent an upper bound to the value of further research as a whole and in relation to the individual parameter groups as they represent the value of perfect information (i.e. assuming that further research will completely resolve any remaining uncertainties). Clearly further research will only partially resolve these uncertainties; the costs reported here were thus only indicative as to the potential value of further research. In effect, they represented a necessary but not sufficient condition for further research to be considered efficient (assuming that the costs of research were actually lower than that represented by the EVPI estimates). Second, these estimates were based on the assumptions and strategies considered in the associated decision model. The model itself focused on strategies for which there was considered sufficiently robust data for them to be included in the evaluation. As such, the full range of potentially feasible strategies was not considered (e.g. although a number of separate cut points was considered for the EPDS, only a single cut point was considered for the BDI). Hence, the finding that the majority of the value associated with the diagnostic performance of the alternative identification strategies was attributed to the use of the EPDS (and not the BDI/Whooley questions) needs to be weighed against the more restrictive range of cut points considered in the case of the BDI and the validity of the diagnostic performance data based on the Whooley questions. Hence, there may still be considerable value that could be

associated with obtaining more reliable data from a range of alternative diagnostic approaches not considered as part of the decision modelling work.

Future research recommendations

Evidence was lacking regarding the effectiveness of methods to identify PND in improving maternal and infant outcomes. Before undertaking a trial, further research would be desirable to determine which, out of the set of potential identification strategies highlighted in this project, is the optimal method to use:

- Evidence to underpin the validity of the two case-finding questions plus the help question in a postnatal population. This could be achieved by undertaking a validation study comparing the diagnostic performances of the EPDS, a generic depression measure (e.g. the BDI) and the two case-finding questions plus the help question with a standardised diagnostic interview conducted according to internationally recognised criteria.
- Evidence to underpin the acceptability of the EPDS, a generic depression measure and the two case-finding questions plus the help question. This could be achieved by conducting a survey on a large sample of women and health-care professionals that asks broad questions about how comfortable respondents feel about using the identification strategy, and exploring this issue in more detail

using qualitative methods such as in-depth interviews. In particular, emphasis should be placed on collating acceptability data by whether women were correctly classified (i.e. true positives or true negatives) or not (i.e. false positives or false negatives). A 2×2 table of acceptability responses could be created.
- Evidence to underpin the natural history of PND over time. This could be achieved by undertaking a longitudinal study with particular emphasis on exploring the population of women who are formally assessed with identification strategies and the population of women in whom formal identification strategies are not utilised.
- Findings from the decision model and the VOI highlight that further evidence is desirable to underpin the costs associated with false positives, the diagnostic performance of identification strategies (primarily related to the use of the EPDS), treatment strategies for confirmed PND, the impact of PND on HRQoL and epidemiological data considered in the model (e.g. prevalence rates).
- Evidence to underpin the clinical effectiveness of the most valid and acceptable instrument to identify PND, identified from the previous two studies,. This could be achieved by undertaking an RCT of an identification strategy with additional training for health-care professionals on the procedures of using the identification strategy versus routine care. Maternal and infant outcomes could be assessed at various time points during the first postnatal year.

Acknowledgments

Contribution of authors

All authors were involved in the conception and design of the study, or the acquisition of data or analysis and interpretation of the data; drafting and revising the report; and final approval of the version to be published.

Individual contributions were as follows: CE Hewitt (Research Fellow) was responsible for the day-to-day activity of the project, study selection, data extraction, conducting the review of validation studies, conducting the review of clinical effectiveness studies, conducting the review of cost-effectiveness studies, quality assessment, data analysis, developing the stakeholder consultation, drafting the full report and making revisions to the report; SM Gilbody (Professor of Psychological Medicine and Health Services Research) was the principal investigator and was involved in the conception and design of the study, day-to-day supervision of the project, developing and implementing the stakeholder consultation and revising the manuscript critically for important intellectual content; S Brealey (Research Fellow) was involved in the day-to-day activity of the project, study selection, data extraction, developing and implementing the stakeholder consultation, drafting and making revisions to the acceptability chapter and commenting on the full report; M Paulden (Research Fellow) was responsible for the economic analysis, the interpretation and presentation of the economic data, drafting and making revisions to the economic chapter and commenting on the report. S Palmer (Senior Research Fellow) was involved in the conception and design of the study, with particular emphasis on the economic component, providing day-to-day supervision of the economic analysis, drafting and revising the economic chapter and commenting on the full report. R Mann (Research Fellow) was involved with quality assessment of all studies in review 1, carrying out the survey of methods to identify PND, drafting of the survey chapter and commenting on the report. J Green (Professor of Psychosocial Reproductive Health and Deputy Director of the Mother and Infant Research Unit) was involved in the conception and design of the study, developing and implementing the stakeholder consultation, interpretation and commenting on the report. J Morrell (Senior Research Fellow) was involved in the conception and design of the study, developing and implementing the stakeholder consultation, interpretation and commenting on the report. M Barkham (Professor of Clinical and Counselling Psychology) was involved in the conception and design of the study, interpretation and commenting on the report. K Light (Information Officer) was involved in devising the search strategy, carrying out the literature searches and writing the search methodology sections of the report. D Richards (Professor of Mental Health) was involved in the conception and design of the study, interpretation and commenting on the report.

References

1. Murray C, Lopez A. *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990*. Boston, MA: Harvard School of Public Health on behalf of the World Bank; 1996.

2. Boath EH, Pryce AJ, Cox JL. Postnatal depression: the impact on the family. *J Reprod Infant Psychol* 1998;**16**:199–203.

3. Lovestone S, Kumar R. Postnatal psychiatric illness: the impact on partners. *Br J Psychiatry* 1993;**163**:210–6.

4. Murray L, Fiori-Cowley A, Hooper R, Cooper P. The impact of postnatal depression and associated adversity on early mother–infant interactions and later infant outcome. *Child Dev* 1996;**67**:2512–26.

5. Murray L, Sinclair D, Cooper P, Ducournau P, Turner P, Stein A. The socioemotional development of 5-year-old children of postnatally depressed mothers. *J Child Psychol Psychiatry* 1999;**40**:1259–71.

6. Cogill S, Caplan H, Alexandra H, Robson K, Kumar R. Impact of maternal postnatal depression on cognitive development of young children. *BMJ* 1986;**292**:1165–7.

7. Hearn G, Iliff A, Jones I, Kirby A, Ormiston P, Parr P, *et al*. Postnatal depression in the community. *Br J Gen Pract* 1998;**48**:1064–6.

8. Henshaw C, Elliott S. *Screening for perinatal depression*. London: Jessica Kingsley; 2005.

9. Cox JL, Murray D, Chapman G. A controlled study of the onset, duration and prevalence of postnatal depression. *Br J Psychiatry* 1993;**163**:27–31.

10. O'Hara MW, Swain AM. Rates and risk of postpartum depression – a meta-analysis. *Int Rev Psychiatry* 1996;**8**:37–54.

11. Cooper PJ, Campbell EA, Day A, Kennerley H, Bond A. Non-psychotic psychiatric disorder after childbirth. A prospective study of prevalence, incidence, course and nature. *Br J Psychiatry* 1988;**152**:799–806.

12. Kumar R, Robson KM. A prospective study of emotional disorders in childbearing women. *Br J Psychiatry* 1984;**144**:35–47.

13. Chisholm D, Diehr P, Knapp M, Patrick D, Treglia M, Simon G. Depression status, medical comorbidity and resource costs. Evidence from an international study of major depression in primary care (LIDO). *Br J Psychiatry* 2003;**183**:121–31.

14. Campbell SB, Cohn JF, Meyers T. Depression in first-time mothers: mother–infant interaction and depression chronicity. *Dev Psychol* 1995;**31**:349–57.

15. Warrington S, Wright C, Team A. Accidents and resulting injuries in premobile infants: data from the ALSPAC study. *Arch Dis Child* 2001;**85**:104–7.

16. Cox J, Holden J, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987;**150**:782–6.

17. Beck CT, Gable RK. Postpartum Depression Screening Scale: development and psychometric testing. *Nurs Res* 2000;**49**:272–82.

18. Beck AT, Ward CH, Mendelson M. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;**4**:561–71.

19. Murray D, Cox JL. Screening for depression during pregnancy with the Edinburgh Depression Scale (EPDS). *J Reprod Infant Psychol* 1990;**8**:99–107.

20. Austin MP, Lumley J. Antenatal screening for postnatal depression: a systematic review. *Acta Psychiatr Scand* 2003;**107**:10–17.

21. Gerrard J, Holden JM, Elliott SA, McKenzie P, McKenzie J, Cox JL. A trainer's perspective of an innovative programme teaching health visitors about the detection, treatment and prevention of postnatal depression. *J Adv Nurs* 1993;**18**:1825–32.

22. Scottish Intercollegiate Guidelines Network. *Management of postnatal depression and puerperal psychosis*. Edinburgh: SIGN Executive; 2002.

23. Secretary of State for Health. *National service framework – mental health*. London: HMSO; 1999.

24. Mant D, Fowler G. Mass screening: theory and ethics. *BMJ* 1990;**300**:916–18.

25. National Screening Committee. *The UK National Screening Committee's criteria for appraising the*

viability, effectiveness and appropriateness of a screening programme. London: HMSO; 2003.

26. Shakespeare J. Screening: the role and recommendations of the UK National Screening Committee. In Henshaw C, Elliott S, editors. *Screening for perinatal depression*. London: Jessica Kingsley; 2005. pp. 21–33.

27. Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G. *et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes*. Evidence Report/Technology Assessment (summary). Maryland: Agency for Healthcare Research and Quality; 2005. pp. 1–8.

28. Pope S, Watts J, Evans S, McDonald S, Henderson J. *Post-natal depression – a systematic review of published scientific literature to 1999*. Canberra: National Health and Medical Research Council; 2000.

29. Dennis C. *Detection, prevention, and treatment of postpartum depression*. 2003. URL: www.who.int/mental_health/prevention/suicide/mmh&chd_exec_sum.pdf. Accessed 20 March 2008.

30. NICE. *Antenatal and postnatal mental health. The NICE guideline on clinical management and service guidance*. London: British Psychological Society and Royal College of Psychiatrists; 2007.

31. Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. *Nurs Res* 2001;**50**:242–50.

32. Boyce PM, Stubbs J, Todd AL. The Edinburgh Postnatal Depression Scale: validation for an Australian sample. *Aust N Z J Psychiatry* 1993;**27**:472–6.

33. Harris B, Huckle P, Thomas R, Johns S, Fung H. The use of rating scales to identify post-natal depression. *Br J Psychiatry* 1989;**154**:813–17.

34. Leverton TJ, Elliott SA. Is the EPDS a magic wand? 1. A comparison of the Edinburgh Postnatal Depression Scale and health visitor report as predictors of diagnosis on the Present State Examination. *J Reprod Infant Psychol* 2000;**18**:279–96.

35. Matthey S, Barnett B, Kavanagh DJ, Howie P. Validation of the Edinburgh Postnatal Depression Scale for men, and comparison of item endorsement with their partners. *J Affect Disord* 2001;**64**:175–84.

36. Murray L, Carothers AD. The validation of the Edinburgh Postnatal Depression Scale on a community sample. *Br J Psychiatry* 1990;**157**:288–90.

37. Zelkowitz P, Milet TH. Screening for post-partum depression in a community sample. *Can J Psychiatry* 1995;**40**:80–6.

38. Whooley MA, Avins AL, Miranda J, Brownwer WS. Case-finding instruments for depression: two questions are as good as many. *J Gen Intern Med* 1997;**12**:439–45.

39. Arroll B, Goodyear-Smith F, Kerse N, Fishman T, Gunn J. Effect of the addition of a 'help' question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. *BMJ* 2005;**331**:884.

40. NHS Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness*. CRD Report No. 4. 2nd edn. York: University of York; 2001.

41. Whiting P, Rutjes A, Reitsma J, Glas A, Bossuyt P, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;**140**:189–202.

42. NHS Centre for Reviews and Dissemination. *Making cost-effectiveness information available: the NHS Economic Evaluation Database Project*. CRD Report No. 6. York: University of York; 1996.

43. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002;**360**:711–15.

44. Beck CT, Gable RK. Further validation of the Postpartum Depression Screening Scale. *Nurs Res* 2001;**50**:155–64.

45. Gruenberg AM, Reed D, Pincus HA. Classification of depression: research and diagnostic criteria: DSM-IV and ICD-10. In Licinio J, Wong ML, editors. *Biology of depression from novel insights to therapeutic strategies*. Weinheim: Wiley-VCH; 2005.

46. American Psychiatric Association. *Diagnostic and statistical manual: mental disorders*. Washington, DC: American Psychiatric Association; 1952.

47. World Health Organization. *Manual of the international classification of diseases, injuries and causes of death, sixth revision*. Geneva: World Health Organization; 1948.

48. American Psychiatric Association. *Diagnostic and statistical manual: mental disorders*. 2nd edition. Washington, DC: American Psychiatric Association; 1968.

49. World Health Organization. *International classification of diseases: manual of the international classification of diseases, injuries and causes of death,*

eighth revision. Geneva: World Health Organization; 1965.

50. Spitzer RL, Endicott J, Robins E. Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry* 1978;**35**:773–82.

51. Feighner JP, Robins SB, Guze RA. Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry* 1972;**26**:57–63.

52. American Psychiatric Association. *Diagnostic and statistical manual: mental disorders*. 3rd edn. Washington, DC: American Psychiatric Association; 1980.

53. World Health Organization. *International classification of diseases, ninth revision*. Geneva: World Health Organization; 1975.

54. World Health Organization. *International classification of diseases, ninth revision, clinical modification*. Ann Arbour, MI: Commission on Professional and Hospital Activities; 1978.

55. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th edn, text revision. Washington, DC: American Psychiatric Association; 2000.

56. World Health Organization. *ICD-10: international statistical classification of diseases and related health problems: tenth revision*. 2nd edn. Geneva: World Health Organization; 2004.

57. Kendell R. Diagnosis and classification of mood disorders. *Psychiatry* 2006;**5**:112–14.

58. Paykal ES. Mood disorders: review of our current diagnostic systems. *Psychopathology* 2002;**35**:94–9.

59. Ruiz P. A research agenda for DSM-V. *Am J Psychiatry* 2004;**161**:1726–7.

60. Elliott SA. Report on the Satra Bruk workshop on classification of post-natal mental disorders. *Arch Women's Ment Health* 2000;**3**:27–33.

61. Whiffen VE. Is postpartum depression a distinct diagnosis? *Clin Psychol Rev* 1992;**12**:485–508.

62. Rogers R. *Handbook of diagnostic and structured interviewing*. New York: Guilford Press; 2001.

63. Stern TA, Herman JB, Slavin PL. *Massachusetts General Hospital guide to primary care psychiatry*. New York, NY: McGraw-Hill Professional; 2003.

64. Weiner IB, Freedheim DK, Velicer WF, Schinka JA, Lerner RM. *Handbook of psychology*. New York, NY: John Wiley; 2003.

65. Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry* 1978;**35**:837–44.

66. Spitzer RL, Williams JBW, Gibbon M. *Users guide for the structured clinical interview for DSM-111-R*. Washington, DC: American Psychiatric Press; 1990.

67. Wing TK, Cooper JE, Sartorius N. *Present State Examination*. Cambridge: Cambridge University Press; 1974.

68. Goldberg D, Cooper B, Eastwood MR. A standardised psychiatric interview for use in community surveys. *Br J Prev Soc Med* 1970;**24**:18–23.

69. Lewis G, Pelosi AJ, Araya RC, Dunn G. Measuring psychiatric disorder in the community: a standardised instrument for use by lay interviewers. *Psychol Med* 1992;**22**:465–86.

70. Robins LN, Helzer JE, Croughan J, Ratcliff KS. The NIMH Diagnostic Interview Schedule, its history, characteristics and validity. *Arch Gen Psychiatry* 1981;**38**:381–9.

71. World Health Organization. *Composite International Diagnostic Interview – version 1.1*. Geneva: WHO; 1993.

72. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, *et al*. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;**59**:22–33.

73. Sheehan DV, Lecrubier Y, Sheehan KH, Janavs J, Weiller E, Keskiner A, *et al*. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *Eur Psychiatry* 1997;**12**:232–41.

74. Lecrubier Y, Sheehan DV, Weiller E, Amorim P, Bonora I, Sheehan KH, *et al*. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry* 1997;**12**:224–31.

75. Raskin A, Schulterbrandt J, Reatig N, McKeon JJ. Replication of factors of psychopathology in interview, ward behaviour and self-report ratings of hospitalized depressives. *J Nerv Ment Dis* 1969;**148**:87–98.

76. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;**134**:382–9.

137

77. Hamilton M. A rating scale for depression. *Neurol Neurosurg Psychiatry* 1960;**23**:56–62.

78. Snaith RP, Constantopoulos AA, Jardine, MY, McGuffin P. A clinical scale for the self-assessment of irritability. *Br J Psychiatry* 1978;**132**:164–71.

79. Cox JL. *Postnatal depression: a guide for health professionals*. Edinburgh: Churchill Livingstone; 1986.

80. Stein G, Van Den Akker O. The retrospective diagnosis of postnatal depression by questionnaire. *J Psychosom Res* 1992;**36**:67–75.

81. Pitt B. 'Atypical' depression following childbirth. *Br J Psychiatry* 1968;**114**:1325–35.

82. Murray L. The impact of postnatal depression on infant development. *J Child Psychol Psychiatry* 1992;**33**:543–61.

83. Williams J, Pignone M, Ramirez G, Perez Stellato C. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry* 2002;**24**:225–37.

84. Goldberg DP. *The detection of psychiatric illness by questionnaire*. London: Oxford University Press; 1972.

85. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;**67**:361–70.

86. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci* 1974;**19**:1–15.

87. Derogatis LR. *SCL-90-R: administration, scoring and procedures manual-II for the R(evised) version and other instruments of the psychopathology rating scale series*. Towson, MD: Clinical Psychometric Research; 1992.

88. Zung WWK. A self-rating depression scale. *Arch Gen Psychiatry* 1965;**12**:63–70.

89. NICE. *Depression: management of depression in primary and secondary care*. London: National Institute for Clinical Excellence; 2004.

90. Spitzer RL, Williams JB, Kroenke K, Linzer M, de Gruy FV III, Hahn SR, *et al*. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA* 1994;**272**:1749–56.

91. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;**3**:32–5.

92. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;**3**:25.

93. Altman DG. *Practical statistics for medical research*. Chapman & Hall/CRC; 1991.

94. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006;**59**:1331–2.

95. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**:982–90.

96. Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *J Clin Epidemiol* 2004;**57**:925–32.

97. Higgins J, Thompson S. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–58.

98. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–60.

99. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol Assess* 2005;**9**(12).

100. StataCorp. Stata Statistical Software: release 9.2; 2007.

101. Harbord R. *metandi: Stata module to perform meta-analysis of diagnostic accuracy*. Bristol: MRC Health Services Research Collaboration, University of Bristol; 2006.

102. Rabe-Hesketh S, Skrondal A, Pickles A. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *J Econom* 2005;**128**:301–23.

103. Terren AC, Garcia Steve L, Navarro P, Aguado J, Ojuel J, Tarragona MJ. Prevalence of postpartum depression in Spanish mothers: comparison of estimation by mean of the structured clinical interview for DSM-IV with the Edinburgh Postnatal Depression Scale. [Spanish]. *Med Clin* 2003;**120**:326–9.

104. Vega-Dienstmaier JM, Suarez GM, Sanchez MC. Validation of a Spanish version of the Edinburgh Postnatal Depression Scale. [Spanish]. *Actas Esp Psiquiatr* 2002;**30**:106–11.

105. Bergant AM, Nguyen T, Heim K, Ulmer H, Dapunt O. German language version and validation of the Edinburgh postnatal depression scale. [German]. *Dtsch Med Wochenschr* 1998;**123**:35–40.

106. Okano T, Murata M, Masuji F, Tamaki R, Nomura J, Miyaoka H, *et al*. Validation and reliability of Japanese version of the EPDS. [Japanese]. *Arch Psychiatr Diagn Clin Eval* 1996;**7**:525–33.

107. Areias MEG, Kumar R, Barros H, Figueiredo E. Comparative incidence of depression in women and men, during pregnancy and after childbirth validation of the Edinburgh postnatal depression scale in Portuguese mothers. *Br J Psychiatry* 1996;**169**:30–5.

108. Abiodun OA. A validity study of the Hospital Anxiety and Depression Scale in general-hospital units and a community sample in Nigeria. *Br J Psychiatry* 1994;**165**:669–72.

109. Adewuya AO, Ola BA, Dada AO, Fasoto OO. Validation of the Edinburgh postnatal depression scale as a screening tool for depression in late pregnancy among Nigerian women. *J Psychosom Obstet Gynecol* 2006;**27**:267–72.

110. Cooper PJ, Murray L, Hooper R, West A. The development and validation of a predictive index for postpartum depression. *Psychol Med* 1996;**26**:627–34.

111. Adouard F, Glangeaud-Freudenthal NM, Golse B. Validation of the Edinburgh postnatal depression scale (EPDS) in a sample of women with high-risk pregnancies in France. *Arch Women's Ment Health* 2005;**8**:89–95.

112. Austin M, Hadzi-Pavlovic D, Saint K, Parker G. Antenatal screening for the prediction of postnatal depression: Validation of a psychosocial Pregnancy Risk Questionnaire. *Acta Psychiatr Scand* 2005;**112**:310–7.

113. Holcomb WL, Jr, Stone LS, Lustman PJ, Gavard JA, Mostello DJ. Screening for depression in pregnancy: characteristics of the Beck Depression Inventory. *Obstet Gynecol* 1996;**88**:1021–5.

114. Kaaya SF, Fawzi MC, Mbwambo JK, Lee B, Msamanga GI, Fawzi W. Validity of the Hopkins Symptom Checklist-25 amongst HIV-positive pregnant women in Tanzania. *Acta Psychiatr Scand* 2002;**106**:9–19.

115. Kitamura T, Shima S, Sugawara M, Toda MA. Temporal variation of validity of self-rating questionnaires: repeated use of the General Health Questionnaire and Zung's Self-rating Depression Scale among women during antenatal and postnatal periods. *Acta Psychiatr Scand* 1994;**90**:446–50.

116. Felice E, Saliba J, Grech V, Cox J. Validation of the Maltese version of the Edinburgh Postnatal Depression Scale. *Arch Women's Ment Health* 2006;**9**:75–80.

117. Werrett J, Clifford C. Validation of the Punjabi version of the Edinburgh postnatal depression scale (EPDS). *Int J Nurs Stud* 2006;**43**:227–36.

118. Thompson WM, Harris B, Lazarus J, Richards C. A comparison of the performance of rating scales used in the diagnosis of postnatal depression. *Acta Psychiatr Scand* 1998;**98**:224–7.

119. Yoshida K, Marks MN, Kibe N, Kumar R, Nakano H, Tashiro N. Postnatal depression in Japanese women who have given birth in England. *J Affect Disord* 1997;**43**:69–77.

120. Cox JL, Chapman G, Murray D, Jones P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. *J Affect Disord* 1996;**39**:185–9.

121. Milgrom J, Ericksen J, Negri L, Gemmill AW. Screening for postnatal depression in routine primary care: properties of the Edinburgh Postnatal Depression Scale in an Australian sample. *Aust N Z J Psychiatry* 2005;**39**:833–9.

122. Barnett B, Matthey S, Gyaneshwar R. Screening for postnatal depression in women of non-English speaking background. *Arch Women's Ment Health* 1999;**2**:67–74.

123. Whiffen VE. Screening for postpartum depression: a methodological note. *J Clin Psychol* 1988;**44**:367–71.

124. Beattie-Clarke P. *Validation of two postpartum screening scales in a sample of Saskatchewan First Nations and Métis women*. Master's thesis. Regina, Canada: University of Regina; 2003.

125. Yamashita H, Yoshida K, Nakano H, Tashiro N. Postnatal depression in Japanese women. Detecting the early onset of postnatal depression by closely monitoring the postpartum mood. *J Affect Disord* 2000;**58**:145–54.

126. Abiodun OA. Postnatal depression in primary care populations in Nigeria. *Gen Hosp Psychiatry* 2006;**28**:133–6.

127. Uwakwe R, Okonkwo JE. Affective (depressive) morbidity in puerperal Nigerian women: validation of the Edinburgh Postnatal Depression Scale. *Acta Psychiatr Scand* 2003;**107**:251–9.

128. Adewuya AO, Eegunranti AB, Lawal AM. Prevalence of postnatal depression in Western Nigerian women: a controlled study. *Int J Psychiatry Clin Pract* 2005;**9**:60–4.

129. Garcia-Esteve L, Ascaso C, Ojuel J, Navarro P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. *J Affect Disord* 2003;**75**:71–6.

130. Navarro P, Ascaso C, Garcia-Esteve L, Aguado J, Torres A, Martin-Santos R. Postnatal psychiatric morbidity: a validation study of the GHQ-12 and the EPDS as screening tools. *Gen Hosp Psychiatry* 2007;**29**:1–7.

131. Muzik M, Klier CM, Rosenblum KL, Holzinger A, Umek W, Katschnig H. Are commonly used self-report inventories suitable for screening postpartum depression and anxiety disorders? *Acta Psychiatr Scand* 2000;**102**:71–3.

132. Jardri R, Pelta J, Maron M, Thomas P, Delion P, Codaccioni X, *et al*. Predictive validation study of the Edinburgh Postnatal Depression Scale in the first week after delivery and risk analysis for postnatal depression. *J Affect Disord* 2006;**93**:169–76.

133. Guedeney N, Fermanian J. Validation study of the French version of the Edinburgh Postnatal Depression Scale (EPDS): new results about use and psychometric properties. *Eur Psychiatry* 1998;**13**:83–9.

134. Benvenuti P, Ferrara M, Niccolai C, Valoriani V, Cox JL. The Edinburgh Postnatal Depression Scale: validation for an Italian sample. *J Affect Disord* 1999;**53**:137–41.

135. Carpiniello B, Pariante CM, Serri F, Costa G, Carta MG. Validation of the Edinburgh Postnatal Depression Scale in Italy. *J Psychosom Obstet Gynecol* 1997;**18**:280–5.

136. Berle JO, Aarre TF, Mykletun A, Dahl AA, Holsten F. Screening for postnatal depression. Validation of the Norwegian version of the Edinburgh Postnatal Depression Scale, and assessment of risk factors for postnatal depression. *J Affect Disord* 2003;**76**:151–6.

137. Eberhard-Gran M, Eskild A, Tambs K, Schei B, Opjordsmoen S. The Edinburgh Postnatal Depression Scale: validation in a Norwegian community sample. *Nord J Psychiatry* 2001;**55**:113–7.

138. Bagedahl-Strindlund M, Monsen Borjesson K. Postnatal depression: a hidden illness. *Acta Psychiatr Scand* 1998;**98**:272–5.

139. Wickberg B, Hwang C. The Edinburgh Postnatal Depression Scale: validation on a Swedish community sample. *Acta Psychiatr Scand* 1996;**94**:181–4.

140. Vittayanont A, Liabsuetrakul T, Pitanupong J. Development of Postpartum Depression Screening Scale (PDSS): a Thai version for screening postpartum depression. *J Med Assoc Thai* 2006;**89**:1–7.

141. Pitanupong J, Liabsuetrakul T, Vittayanont A. Validation of the Thai Edinburgh Postnatal Depression Scale for screening postpartum depression. *Psychiatry Res* 2007;**149**:253–9.

142. Aydin N, Inandi T, Yigit A, Hodoglugil NNS. Validation of the Turkish version of the Edinburgh Postnatal Depression Scale among women within their first postpartum year. *Soc Psychiatry Psychiatr Epidemiol* 2004;**39**:483–6.

143. Aslan E, Kuey L, Kultur S. The validity and reliability study of the Turkish version of the Edinburgh Postnatal Depression Scale. Spring Symposium First Book. Ankara: Psychiatric Association of Turkey; 1997. pp. 51–2.

144. Beck CT, Gable RK. Screening performance of the postpartum depression screening scale – Spanish version. *J Transcult Nurs* 2005;**16**:331–8.

145. Jadresic E, Araya R, Jara C. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Chilean postpartum women. *J Psychosom Obstet Gynecol* 1995;**16**:187–91.

146. Lee DT, Yip AS, Chiu HF, Leung TY, Chung TK. Screening for postnatal depression: are specific instruments mandatory? *J Affect Disord* 2001;**63**:233–8.

147. Lee DT, Yip AS, Chiu HF, Chung TK. Screening for postnatal depression using the double-test strategy. *Psychosom Med* 2000;**62**:258–63.

148. Lee DTS, Yip ASK, Chan SSM, Tsui MHY, Wong WS, Chung TKH. Postdelivery screening for postpartum depression. *Psychosom Med* 2003;**65**:357–61.

149. Mahmud WMR, Awang A, Herman I, Mohamed MN. Analysis of the psychometric properties of the Malay version of Beck Depression Inventory II (BDI-II) among postpartum women in Kedah, North West of Peninsular Malaysia. *Malays J Med Sci* 2004;**11**:19–25.

150. Mahmud W, Awang A, Mohamed MN. Revalidation of the Malay version of the Edinburgh postnatal depression scale (EPDS) among Malay postnatal women attending the Bakar Bata Health Center in Alor Setar, Kedah, North West of Peninsular Malaysia. *Malays J Med Sci* 2003;**10**:71–5.

151. Agoub M, Moussaoui D, Battas O. Prevalence of postpartum depression in a Moroccan sample. *Arch Women's Ment Health* 2005;**8**:37–43.

152. Regmi S, Sligl W, Carter D, Grut W, Seear M. A controlled study of postpartum depression among Nepalese women: validation of the Edinburgh Postpartum Depression Scale in Kathmandu. *Trop Med Int Health* 2002;**7**:378–82.

153. Holt WJ. The detection of postnatal depression in general practice using the Edinburgh postnatal depression scale. *N Z Med J* 1995;**108**:57–9.

154. Lawrie TA, Hofmeyr GJ, De Jager M, Berk M. Validation of the Edinburgh Postnatal Depression Scale on a cohort of South African women. *S Afr Med J* 1998;**88**:1340–4.

155. Teng H-W, Hsu C-S, Shih S-M, Lu M-L, Pan J-J, Shen WW. Screening postpartum depression with the Taiwanese version of the Edinburgh Postnatal Depression Scale. *Compre Psychiatry* 2005;**46**:261–5.

156. Ghubash R, Abou-Saleh MT, Daradkeh TK. The validity of the Arabic Edinburgh Postnatal Depression Scale. *Soc Psychiatry Psychiatr Epidemiol* 1997;**32**:474–6.

157. Smith LK, Pope C, Botha JL. Patients' help-seeking experiences and delay in cancer presentation: a qualitative synthesis. *Lancet* 2005;**366**:825–31.

158. Noyes J, Popay J. Directly observed therapy and tuberculosis: how can a systematic review of qualitative research contribute to improving services? A qualitative meta-synthesis. *J Adv Nurs* 2007;**57**:227–43.

159. Thomas J, Harden A, Oakley A, *et al*. Integrating qualitative research with trials in systematic reviews. *BMJ* 2004;**328**:1010–12.

160. Boulton M, Fitzpatrick R, Swinburn C. Qualitative research in health care. II. A structured review and evaluation of studies. *J Eval Clin Pract* 1996;**2**:171–9.

161. Mays N, Pope C. Qualitative research: rigour and qualitative research. *BMJ* 1995;**311**:109.

162. Crombie I. *The pocket guide to critical appraisal*. London: BMJ Publishing; 1996.

163. Petticrew M, Roberts H. *Systematic reviews in the social sciences. A practice guide*. Oxford: Blackwell Publishing; 2006.

164. Cochrane Qualitative Methods Group. URL: www.joannabriggs.edu.au/cqrmg/tools_4.html.

165. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers, M, *et al*. *Guidance on the conduct of narrative synthesis in systematic reviews*. URL: www.lancs.ac.uk/fass/projects/nssr/.

166. Lucas PJ, Baird J, Arai L, Law C, Roberts HM. Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews. *BMC Med Res Methodol* 2007;**7**:4.

167. Shakespeare J, Blake F, Garcia J. A qualitative study of the acceptability of routine screening of postnatal women using the Edinburgh Postnatal Depression Scale. *Br J Gen Pract* 2003;**53**:614–9.

168. Poole H, Mason L, Osborn T. Women's views of being screened for postnatal depression. *Community Pract* 2006;**79**:363–7.

169. Cubison J, Munro J. Acceptability of using the EPDS as a screening tool for postnatal depression. In Henshaw C, Elliott S, editors. *Screening for perinatal depression*. London: Jessica Kingsley; 2005. pp. 152–61.

170. Matthey S, Barnett BEW, Elliott A. Vietnamese and Arabic women's responses to the Diagnostic Interview Schedule (depression) and self-report questionnaires: cause for concern. *Aust N Z J Psychiatry* 1997;**31**:360–9.

171. Gemmill AW, Leigh B, Ericksen J, Milgrom J. A survey of the clinical acceptability of screening for postnatal depression in depressed and non-depressed women. *BMC Public Health* 2006;**6**:211.

172. Poole H, Mason L. Health visitors views of screening for post natal depression. Liverpool: John Moore University; 2008.

173. Brown H, Bacigalupo R. Health visitors and postnatal depression: identification and practice. *Community Pract* 2006;**79**:49–52.

174. Matthey S, White T, Phillips J, Taouk R, Chee TT, Barnett B. Acceptability of routine antenatal psychosocial assessments to women from English and non-English speaking backgrounds. *Arch Womens Ment Health* 2005;**8**:171–80.

175. Clark G. Discussing emotional health in pregnancy: the Edinburgh Postnatal Depression Scale. *Br J Community Nurs* 2000;**5**:91–8.

176. Jacquemain F, Golse B. Predictability of maternal post-partum depression: creation of a screening questionnaire. *Psichiatria dell'Infancia e dell'Adolescenza*. 1998;**65**:391–406.

177. Wood C. *An investigation of the utility of an antenatal screening index in aiding health visitors to predict postnatal depression*. NHS Lothian; 2002.

178. Edge D. What do black Caribbean women think about screening with the EPDS? In Henshaw C, Elliott S, editors. *Screening for perinatal depression*. London: Jessica Kingsley; 2005. pp. 162–70.

179. Buist A, Condon J, Brooks J, Speelman C, Milgrom J, Hayes B, *et al*. Acceptability of routine screening for perinatal depression. *J Affect Disord* 2006;**93**:233–7.

180. Alder E. NAPD: the national audit of the detection and management of postnatal depression. NHS Scotland; 2007.

181. Elliott SA, Gerrard J, Ashton C, Cox JL. Training health visitors to reduce levels of depression after childbirth: an evaluation. *J Ment Health* 2001;**10**:613–25.

182. Waldenstrom U. Why do some women change their opinion about childbirth over time? *Birth* 2004;**31**:102–7.

183. Hoffbrand S, Howard L, Crawley H. Antidepressant treatment for postnatal depression. *Cochrane Database Syst Rev* 2001;**2**:CD002018.

184. Dennis CL, Creedy D. Psychosocial and psychological interventions for preventing postpartum depression. *Cochrane Database Syst Rev* 2004;**4**:CD001134.

185. Howard LM, Hoffbrand S, Henshaw C, Boath L, Bradley E. Antidepressant prevention of postnatal depression. *Cochrane Database Syst Rev* 2005;**2**:CD4363.

186. Dennis CL, Giavedoni K. Interventions (other than pharmacological, psychosocial or psychological) for treating antenatal depression. *Cochrane Database Syst Rev* 2008;**4**:CD006795.

187. Dennis CL, Hodnett E. Psychosocial and psychological interventions for treating postpartum depression. *Cochrane Database Syst Rev* 2007;**4**:CD006116.

188. Dennis CL, Ross LE, Grigoriadis S. Psychosocial and psychological interventions for treating antenatal depression. *Cochrane Database Syst Rev* 2007;**3**:CD.

189. Pignone MP, Gaynes BN, Rushton JL, Burchell CM, Orleans T, Mulrow CD, *et al*. Screening for depression in adults: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med* 2002;**136**:765–76.

190. Gilbody S, Whitty P, Grimshaw J, Thomas R. Educational and organizational interventions to improve the management of depression in primary care: a systematic review. *JAMA* 2003;**289**:3145–51.

191. Evins GG, Theofrastous JP, Galvin SL. Postpartum depression: a comparison of screening and routine clinical evaluation. *Am J Obstet Gynecol* 2000;**182**:1080–2.

192. Wickberg B, Tjus T, Hwang P. Using the EPDS in routine antenatal care in Sweden: a naturalistic study. *J Reprod Infant Psychol* 2005;**23**:33–41.

193. MacArthur C, Winter HR, Bick DE, Knowles H, Lilford R, Henderson C, *et al*. Effects of redesigned community postnatal care on womens' health 4 months after birth: a cluster randomised controlled trial. *Lancet* 2002;**359**:378–85.

194. Kung F, Lee I, Lee CP. Can midwife help in preventing depression in women at six weeks postpartum by early intervention? In International Confederation of Midwives – Midwives and Women Working Together for the Family of the World: ICM proceedings CD ROM, Vienna, 2002.

195. Webster J, Linnane J, Roberts J, Starrenburg S, Hinson J, Dibley L. IDentify, Educate and Alert (IDEA) trial: an intervention to reduce postnatal depression. *Br J Obstet Gynaecol* 2003;**110**:842–6.

196. Brugha TS, Wheatley S, Taub NA, Culverwell A, Friedman T, Kirwan P, *et al*. Pragmatic randomized trial of antenatal intervention to prevent post-natal depression by reducing psychosocial risk factors. *Psychol Med* 2000;**30**:1273–81.

197. Buist A, Westley D, Hill C. Antenatal prevention of postnatal depression. *Arch Women's Ment Health* 1999;**1**:167–73.

198. Elliott SA, Sanjack M, Leverton TJ. Parents groups in pregnancy: a preventive intervention for postnatal depression? In Gottlieb BH, editor. *Marshaling social support: formats, processes, and effects*. Thousand Oaks, CA: Sage Publications; 1988. pp. 87–110.

199. Elliott SA, Leverton TJ, Sanjack M, Turner H, Cowmeadow P, Hopkins J, *et al*. Promoting mental health after childbirth: a controlled trial of primary prevention of postnatal depression. *Br J Clin Psychol* 2000;**39**:223–41.

200. Gorman L. *Prevention of postpartum depression in a high risk sample*. Iowa City, IA: University of Iowa; 2001.

201. McKee M, Zayas LH, Fletcher J, Boyd RC, Nam SH. Results of an intervention to reduce perinatal depression among low-income minority women in community primary care. *J Soc Serv Res* 2006;**32**:63–81.

202. Zayas LH, McKee MD, Jankowski KRB. Adapting psychosocial intervention research to urban primary care environments: a case example. *Ann Fam Med* 2004;**2**:504–8.

203. Stamp EE, Williams AS, Crowther CA. Evaluation of antenatal and postnatal support to overcome

postnatal depression – a randomized, controlled trial. *Birth* 1995;**22**:138–43.

204. Zlotnick C, Johnson SL, Miller IW, Pearlstein T, Howard M. Postpartum depression in women receiving public assistance: pilot study of an interpersonal-therapy-oriented group intervention. *Am J Psychiatry* 2001;**158**:638–40.

205. Zlotnick C, Miller IW, Pearlstein T, Howard M, Sweeney P. A preventive intervention for pregnant women on public assistance at risk for postpartum depression. *Am J Psychiatry* 2006;**163**:1443–5.

206. Armstrong K, Edwards H. The effects of exercise and social support on mothers reporting depressive symptoms: a pilot randomized controlled trial. *Int J Ment Health Nurs* 2003;**12**:130–8.

207. Dennis C-L. The effect of peer support on postpartum depression: a pilot randomized controlled trial. *Can J Psychiatry* 2003;**48**:115–24.

208. Heh S, Fu Y. Effectiveness of informational support in reducing the severity of postnatal depression in Taiwan. *J Adv Nurs* 2003;**42**:30–6.

209. Honey KL, Bennett P, Morgan M. A brief psycho-educational group intervention for postnatal depression. *Br J Clin Psychol* 2002;**41**:405–9.

210. Ingadottir E, Thome M. Evaluation of a web-based course for community nurses on postpartum emotional distress. *Scand J Caring Sci* 2006;**20**:86–92.

211. Misri S, Reebye P, Corral M, Milis L. The use of paroxetine and cognitive-behavioral therapy in postpartum depression and anxiety: a randomized controlled trial. *J Clin Psychiatry* 2004;**65**:1236–41.

212. Prendergast J, Austin M-P. Early childhood nurse-delivered cognitive behavioural counselling for post-natal depression. *Aust Psychiatry* 2001;**9**:255–9.

213. O'Hara MW, Stuart S, Gorman LL, Wenzel A. Efficacy of interpersonal psychotherapy for postpartum depression. *Arch Gen Psychiatry* 2000;**57**:1039–45.

214. Chen CH, Tseng YF, Chou FH, Wang SY. Effects of support group intervention in postnatally distressed women. A controlled study in Taiwan. *J Psychosom Res* 2000;**49**:395–9.

215. Chabrol H, Teissedre F, Saint-Jean M, Teisseyre N, Roge B, Mullet E. Prevention and treatment of post-partum depression: a controlled randomized study on women at risk. *Psychol Med* 2002;**32**:1039–47.

216. Cooper PJ, Murray L, Wilson A, Romaniuk H. Controlled trial of the short- and long-term effect of

psychological treatment of post-partum depression. 1: impact on maternal mood. *Br J Psychiatry* 2003;**182**:412–9.

217. Murray L, Cooper P, Wilson A, Romaniuk H. Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression. *Br J Psychiatry* 2003;**182**:420–7.

218. Holden JM, Sagovsky R, Cox JL. Counselling in a general practice setting: controlled study of heath visitor intervention in treatment of postnatal depression. *BMJ Clin Res Ed* 1989;**298**:223–6.

219. Wickberg B, Hwang C. Counselling of postnatal depression: a controlled study on a population-based Swedish sample. *J Affect Disord* 1996;**39**:209–16.

220. Meager I, Milgrom J. Group treatment for postpartum depression: a pilot study. *Aust N Z J Psychiatry* 1996;**30**:852–60.

221. Horowitz JA, Bell M, Trybulski J, Munro BH, Moser D, Hartz SA, *et al.* Promoting responsiveness between mothers with depressive symptoms and their infants. *J Nurs Scholarsh* 2001;**33**:323–9.

222. Milgrom J, Negri LM, Gemmill AW, McNeil M, Martin PR. A randomized controlled trial of psychological interventions for postnatal depression. *Br J Clin Psychol* 2005;**44**:529–42.

223. Drummond M. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press; 1997.

224. Morrell CJ, Spiby H, Stewart P, Walters S, Morgan A. Costs and effectiveness of community postnatal support workers: randomised controlled trial. *BMJ* 2000;**321**:593–8.

225. Boath E, Major K, Cox J. When the cradle falls. II: the cost-effectiveness of treating postnatal depression in a psychiatric day hospital compared with routine primary care. *J Affect Disord* 2003;**74**:159–66.

226. Appleby L, Hirst E, Marshall S, Keeling F, Brind J, Butterworth T, *et al.* The treatment of postnatal depression by health visitors: impact of brief training on skills and clinical practice. *J Affect Disord* 2003;**77**:261–6.

227. Petrou S, Cooper P, Murray L, Davidson LL. Cost-effectiveness of a preventive counseling and support package for postnatal depression. *Int J Technol Assess Health Care* 2006;**22**:443–53.

228. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis

using Monte Carlo simulation. *Med Decis Making* 1985;**5**:157–77.

229. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making* 2002;**22**:290–308.

230. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;**18**:342–64.

231. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the National Health Service Health Technology Assessment Programme. *Health Technol Assess* 2004;**8**.

232. Curtis K, Netten A. *Unit costs of health and social care 2005*. Canterbury: Personal Social Services Research Unit, University of Kent; 2005.

233. Curtis K. *Unit costs of health and social care 2006*. Canterbury: Personal Social Services Research Unit, University of Kent; 2007.

234. Revicki D, Wood M. Patient-assigned health state utilities for depression-related outcomes: differences by depression severity and antidepressant medications. *J Affect Disord* 1998;48:25–36.

235. NICE. *Guide to the methods of technology appraisal*. London: NICE; 2004.

236. Kessler D, Bennewith O, Lewis G, Sharp D. Detection of depression and anxiety in primary care: follow up study. *BMJ* 2002;**325**:1016–7.

237. Microsoft Corporation. Excel. Redmond, WA; 2003.

238. Johannesson M. On the estimation of cost-effectiveness ratios. *Health Policy* 1995;**31**:225–9.

239. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001;**10**:779–89.

240. Fenwick E, O'Brien BJ, Briggs AH. Cost-effectiveness acceptability curves – facts, fallacies and FAQ. *Health Econ* 2004;**13**:405–15.

241. Chisholm D, Healey A, Knapp M. QALYs and mental health care. *Soc Psychiatry Psychiatr Epidemiol* 1979;**32**:68–75.

242. Thomas C, Morris S. Cost of depression among adults in England in 2000. *Br J Psychiatry* 2003:**83**:514–9.

243. Office for National Statistics. *Birth statistics: review of the Registrar General on births and patterns of family building in England and Wales, 2006*. Newport: Office for National Statistics; 2007.

244. Baggaley RF, Ganaba R, Filippi V, Kere M, Marshall T, Sombie I, *et al*. Detecting depression after pregnancy: the validity of the K10 and K6 in Burkina Faso. *Trop Med Int Health* 2007;**12**:1225–9.

245. Cantilino A, Carvalho JA, Maia A, Albuquerque C, Cantilino G, Sougey EB. Translation, validation and cultural aspects of postpartum depression screening scale in Brazilian Portuguese. *Transcult Psychiatry* 2007;**44**:672–84.

246. Hanlon C, Medhin G, Alem A, Araya M, Abdulahi A, Hughes M, *et al*. Detecting perinatal common mental disorders in Ethiopia: validation of the self-reporting questionnaire and Edinburgh Postnatal Depression Scale. *J Affect Disord* 2008;**108**:251–62.

247. Hanusa BH, Scholle SH, Haskett RF, Spadaro K, Wisner KL. Screening for depression in the postpartum period: a comparison of three instruments. *J Women's Health* 2008;**17**:585–96.

248. Karmaliani R, Bann CM, Pirani F, Akhtar S, Bender RH, goldenberg RL, *et al*. Diagnostic validity of two instruments for assessing anxiety and depression among pregnant women in Hyderabad, Pakistan. *Health Care Women Int* 2007;**28**:556–72.

249. Mazhari S, Nakhaee N. Validation of the Edinburgh Postnatal Depression Scale in an Iranian sample. *Arch Women's Ment Health* 2007;**10**:293–7.

250. Rowe HJ, Fisher JRW, Loh WM. The Edinburgh Postnatal Depression Scale detects but does not distinguish anxiety disorders from depression in mothers of infants. *Arch Women's Ment Health* 2008;**11**:103–8.

251. Rowel D, Jayawardena P, Fernando N. Validation of the Sinhala translation of Edinburgh Postnatal Depression Scale. *Ceylon Med J* 2008;**53**:10–13.

252. Su K-P, Chiu T-H, Huang C-L, Pariante CM. Different cutoff points for different trimesters? The use of Edinburgh Postnatal Depression Scale and Beck Depression Inventory to screen for depression in pregnant Taiwanese women. *Gen Hosp Psychiatry* 2007;**29**:436–41.

253. Santos IS, Matijasevich A, Tavares BF, Barros AJD, Botello IP, Lapolli C, *et al*. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in a sample of mothers from the 2004 Pelotas Birth Cohort Study. *Cad Saude Publica* 2007;**23**:2577–88.

254. Liabsuetrakul T, Vittayanont A, Pitanupong J. Clinical applications of anxiety, social support, stressors, and self-esteem measured during pregnancy and postpartum for screening postpartum depression in Thai women. *J Obstet Gynaecol Res* 2007;**33**:333–40.

255. Leigh B, Milgrom J. Acceptability of antenatal screening for depression in routine antenatal care. *Aust J Adv Nurs* 2007;**24**:14–18.

256. Massoudi P, Wickberg B, Hwang P. Screening for postnatal depression in Swedish child health care. *Acta Paediatr* 2007;**96**:897–901.

257. Forman DR, O'Hara MW, Stuart S, Gorman LL, Larsen KE, Coy KC. Effective treatment for postpartum depression is not sufficient to improve the developing mother-child relationship. *Dev Psychopathol* 2007;**19**:585–602.

258. Heh S-S, Huang L-H, Ho S-M, Fu Y-Y, Wang L-L. Effectiveness of an exercise support program in reducing the severity of postnatal depression in Taiwanese women. *Birth* 2008;**35**:60–5.

259. Rojas G, Fritsch R, Solis J, Jadresic E, Castillo C, Gonzalez M, *et al*. Treatment of postnatal depression in low-income mothers in primary-care clinics in Santiago, Chile: a randomised controlled trial. *Lancet* 2007;**370**:1629–37.

145

Appendix 1

Full search strategy

MEDLINE (Ovid Online – www.ovid.com/) 1950 to January Week 5 2007

Searched 8 February 2007.
Retrieved 4341 hits.

1. Pregnancy/
2. Prenatal Care/
3. Postnatal Care/
4. pregnancy.ti,ab.
5. pregnant.ti,ab.
6. prenatal.ti,ab.
7. pre-natal.ti,ab.
8. postnatal.ti,ab.
9. postnatal.ti,ab.
10. postpartum.ti,ab.
11. post-partum.ti,ab.
12. puerperal.ti,ab.
13. new mother$.ti,ab.
14. pre-pregnancy.ti,ab.
15. prepregnancy.ti,ab.
16. ante-natal.ti,ab.
17. antenatal.ti,ab.
18. antepartum.ti,ab.
19. ante-partum.ti,ab.
20. or/1–19
21. Depression/
22. Depression, Postpartum/
23. pnd.ti,ab.
24. blues.ti,ab.
25. depress$.ti,ab.
26. Depressive Disorder/
27. melancholia.ti,ab.
28. (anxiety or anxious).ti,ab.
29. anxiety/
30. ppd.ti,ab.
31. or/21–30
32. screen$.ti,ab.
33. diagnos$.ti,ab.
34. detect$.ti,ab.
35. predict$.ti,ab.
36. aware$.ti,ab.
37. identif$.ti,ab.
38. DIAGNOSIS/
39. (edinburgh adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
40. EPDS.ti,ab.
41. (Postpartum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
42. (Post-partum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
43. PDSS.ti,ab.
44. (Bromley adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
45. BPDS.ti,ab.
46. (General Health adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
47. GHQ.ti,ab.
48. (Beck adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
49. BDI.ti,ab.
50. BAI.ti,ab.
51. (State adj2 anxiety adj2 depression).ti,ab.
52. SAD.ti,ab.
53. (Hospital adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
54. HADS.ti,ab.
55. (Hamilton adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
56. HRSD.ti,ab.
57. (Zung adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
58. SDS.ti,ab.
59. Profile of mood states.ti,ab.
60. POMS.ti,ab.
61. (Centre adj2 Epidemiological studies adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
62. (CES-D or CESD).ti,ab.
63. Symptom Checklist-90-revised.ti,ab.
64. SCL-90-R.ti,ab.
65. (Brief symptom adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
66. BSI.ti,ab.
67. ((Inventory or Questionnaire or scale or index or checklist or interview) adj5 depressive symptomatology).ti,ab.
68. IDS.ti,ab.

147

69. (Montgomery-Asberg adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
70. MADRS.ti,ab.
71. (Depressive Adjective adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
72. DACL.ti,ab.
73. (Schedule adj2 affective disorders adj2 schizophrenia).ti,ab.
74. SADS.ti,ab.
75. (State-Trait anxiety adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
76. STAI.ti,ab.
77. (Brisbane postnatal adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
78. (Post-partum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
79. (Postpartum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
80. ((Depress$or anxiety) adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
81. QUESTIONNAIRES/
82. INTERVIEWS/
83. antenatal psychosocial health assessment. ti,ab.
84. alpha.ti,ab.
85. or/32–84
86. 20 and 31 and 85

CINAHL (Cumulative Index to Nursing and Allied Health Literature) (Ovid Online – www.ovid.com/) 1982 to February Week 1 2007

Searched 8 February 2007.
Retrieved 1346 hits.

1. Pregnancy/
2. Prenatal Care/
3. Postnatal Care/
4. pregnancy.ti,ab.
5. pregnant.ti,ab.
6. prenatal.ti,ab.
7. pre-natal.ti,ab.
8. postnatal.ti,ab.
9. postnatal.ti,ab.
10. postpartum.ti,ab.
11. post-partum.ti,ab.
12. puerperal.ti,ab.
13. new mother$.ti,ab.
14. pre-pregnancy.ti,ab.
15. prepregnancy.ti,ab.
16. ante-natal.ti,ab.
17. antenatal.ti,ab.
18. antepartum.ti,ab.
19. ante-partum.ti,ab.
20. or/1–19
21. Depression/
22. Depression, Postpartum/
23. pnd.ti,ab.
24. blues.ti,ab.
25. depress$.ti,ab.
26. melancholia.ti,ab.
27. (anxiety or anxious).ti,ab.
28. anxiety/
29. ppd.ti,ab.
30. or/21–29
31. screen$.ti,ab.
32. diagnos$.ti,ab.
33. detect$.ti,ab.
34. predict$.ti,ab.
35. aware$.ti,ab.
36. identif$.ti,ab.
37. DIAGNOSIS/
38. (edinburgh adj5 (Inventory or Questionnaire or scale or index or checklist or interview)). ti,ab.
39. EPDS.ti,ab.
40. (Postpartum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
41. (Post-partum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
42. PDSS.ti,ab.
43. (Bromley adj5 (Inventory or Questionnaire or scale or index or checklist or interview)). ti,ab.
44. BPDS.ti,ab.
45. (General Health adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
46. GHQ.ti,ab.
47. (Beck adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
48. BDI.ti,ab.
49. BAI.ti,ab.
50. (State adj2 anxiety adj2 depression).ti,ab.
51. SAD.ti,ab.
52. (Hospital adj5 (Inventory or Questionnaire or scale or index or checklist or interview)). ti,ab.
53. HADS.ti,ab.
54. (Hamilton adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.

55. HRSD.ti,ab.
56. (Zung adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
57. SDS.ti,ab.
58. Profile of mood states.ti,ab.
59. POMS.ti,ab.
60. (Centre adj2 Epidemiological studies adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
61. (CES-D or CESD).ti,ab.
62. Symptom Checklist-90-revised.ti,ab.
63. SCL-90-R.ti,ab.
64. (Brief symptom adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
65. BSI.ti,ab.
66. ((Inventory or Questionnaire or scale or index or checklist or interview) adj5 depressive symptomatology).ti,ab.
67. IDS.ti,ab.
68. (Montgomery-Asberg adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
69. MADRS.ti,ab.
70. (Depressive Adjective adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
71. DACL.ti,ab.
72. (Schedule adj2 affective disorders adj2 schizophrenia).ti,ab.
73. SADS.ti,ab.
74. (State-Trait anxiety adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
75. STAI.ti,ab.
76. (Brisbane postnatal adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
77. (Post-partum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
78. (Postpartum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
79. ((Depress$ or anxiety) adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
80. exp QUESTIONNAIRES/
81. exp INTERVIEWS/
82. scales/
83. antenatal psychosocial health assessment. ti,ab.
84. alpha.ti,ab.
85. beck depression inventory, revised edition/
86. beck hopelessness scale/
87. centerforepidemiologicalstudiesdepressionscale/

88. edinburgh postnatal depression scale/
89. hamilton rating scale for depression/
90. "profile of mood states, revised"/
91. psychiatric symptom index/
92. self-rating anxiety scale/
93. self-rating depression scale/
94. state-trait anxiety inventory/
95. brief symptom inventory/
96. or/31–95
97. 20 and 30 and 96

PsycINFO (Ovid Online – www.ovid.com/) 1806 to January Week 5 2007

Searched 8 February 2007.
Retrieved 2255 hits.

1. exp Pregnancy/
2. Prenatal Care/
3. postnatal period/
4. Perinatal Period/
5. pregnancy.ti,ab.
6. pregnant.ti,ab.
7. prenatal.ti,ab.
8. pre-natal.ti,ab.
9. postnatal.ti,ab.
10. postnatal.ti,ab.
11. postpartum.ti,ab.
12. post-partum.ti,ab.
13. puerperal.ti,ab.
14. new mother$.ti,ab.
15. pre-pregnancy.ti,ab.
16. prepregnancy.ti,ab.
17. ante-natal.ti,ab.
18. antenatal.ti,ab.
19. antepartum.ti,ab.
20. ante-partum.ti,ab.
21. or/1–20
22. Major Depression/
23. Postpartum Depression/
24. pnd.ti,ab.
25. blues.ti,ab.
26. depress$.ti,ab.
27. melancholia.ti,ab.
28. (anxiety or anxious).ti,ab.
29. anxiety/
30. ppd.ti,ab.
31. or/22–30
32. screen$.ti,ab.
33. diagnos$.ti,ab.
34. detect$.ti,ab.
35. predict$.ti,ab.
36. aware$.ti,ab.
37. identif$.ti,ab.
38. DIAGNOSIS/

149

39. exp Psychodiagnosis/
40. (edinburgh adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
41. EPDS.ti,ab.
42. (Postpartum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
43. (Post-partum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
44. PDSS.ti,ab.
45. (Bromley adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
46. BPDS.ti,ab.
47. (General Health adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
48. GHQ.ti,ab.
49. (Beck adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
50. BDI.ti,ab.
51. BAI.ti,ab.
52. (State adj2 anxiety adj2 depression).ti,ab.
53. SAD.ti,ab.
54. (Hospital adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
55. HADS.ti,ab.
56. (Hamilton adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
57. HRSD.ti,ab.
58. (Zung adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
59. SDS.ti,ab.
60. Profile of mood states.ti,ab.
61. POMS.ti,ab.
62. (Centre adj2 Epidemiological studies adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
63. (CES-D or CESD).ti,ab.
64. Symptom Checklist-90-revised.ti,ab.
65. SCL-90-R.ti,ab.
66. (Brief symptom adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
67. BSI.ti,ab.
68. ((Inventory or Questionnaire or scale or index or checklist or interview) adj5 depressive symptomatology).ti,ab.
69. IDS.ti,ab.
70. (Montgomery-Asberg adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
71. MADRS.ti,ab.
72. (Depressive Adjective adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
73. DACL.ti,ab.
74. (Schedule adj2 affective disorders adj2 schizophrenia).ti,ab.
75. SADS.ti,ab.
76. (State-Trait anxiety adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
77. STAI.ti,ab.
78. (Brisbane postnatal adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
79. (Post-partum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
80. (Postpartum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
81. ((Depress$or anxiety) adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
82. QUESTIONNAIRES/
83. INTERVIEWS/
84. antenatal psychosocial health assessment.ti,ab.
85. alpha.ti,ab.
86. or/32–84

EMBASE (Ovid Online – www.ovid.com/) 1980 to 2007 Week 5

Searched 8 February 2007.
Retrieved 3736 hits.

1. exp Pregnancy/
2. Prenatal Care/
3. perinatal care/
4. maternal care/
5. puerperium/
6. Postnatal Care/
7. pregnancy.ti,ab.
8. pregnant.ti,ab.
9. prenatal.ti,ab.
10. pre-natal.ti,ab.
11. postnatal.ti,ab.
12. postnatal.ti,ab.
13. postpartum.ti,ab.
14. post-partum.ti,ab.
15. puerperal.ti,ab.
16. new mother$.ti,ab.
17. pre-pregnancy.ti,ab.
18. prepregnancy.ti,ab.
19. ante-natal.ti,ab.

20. antenatal.ti,ab.
21. antepartum.ti,ab.
22. ante-partum.ti,ab.
23. or/1–22
24. Depression/
25. major depression/
26. melancholia/
27. Puerperal Depression/
28. pnd.ti,ab.
29. blues.ti,ab.
30. depress$.ti,ab.
31. Depressive Disorder/
32. melancholia.ti,ab.
33. (anxiety or anxious).ti,ab.
34. "mixed anxiety and depression"/
35. ppd.ti,ab.
36. or/24–35
37. screen$.ti,ab.
38. Screening/
39. diagnos$.ti,ab.
40. detect$.ti,ab.
41. predict$.ti,ab.
42. aware$.ti,ab.
43. identif$.ti,ab.
44. DIAGNOSIS/
45. (edinburgh adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
46. EPDS.ti,ab.
47. (Postpartum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
48. (Post-partum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
49. PDSS.ti,ab.
50. (Bromley adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
51. BPDS.ti,ab.
52. (General Health adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
53. GHQ.ti,ab.
54. (Beck adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
55. BDI.ti,ab.
56. BAI.ti,ab.
57. (State adj2 anxiety adj2 depression).ti,ab.
58. SAD.ti,ab.
59. (Hospital adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
60. HADS.ti,ab.
61. (Hamilton adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
62. HRSD.ti,ab.
63. (Zung adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
64. SDS.ti,ab.
65. Profile of mood states.ti,ab.
66. POMS.ti,ab.
67. (Centre adj2 Epidemiological studies adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
68. (CES-D or CESD).ti,ab.
69. Symptom Checklist-90-revised.ti,ab.
70. SCL-90-R.ti,ab.
71. (Brief symptom adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
72. BSI.ti,ab.
73. ((Inventory or Questionnaire or scale or index or checklist or interview) adj5 depressive symptomatology).ti,ab.
74. IDS.ti,ab.
75. (Montgomery-Asberg adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
76. MADRS.ti,ab.
77. (Depressive Adjective adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
78. DACL.ti,ab.
79. (Schedule adj2 affective disorders adj2 schizophrenia).ti,ab.
80. SADS.ti,ab.
81. (State-Trait anxiety adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
82. STAI.ti,ab.
83. (Brisbane postnatal adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
84. (Post-partum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
85. (Postpartum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
86. ((Depress$ or anxiety) adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
87. exp Questionnaire/
88. exp Interview/
89. antenatal psychosocial health assessment.ti,ab.
90. alpha.ti,ab.
91. or/37–90
92. 23 and 36 and 91

**Maternity and Infant Care
(Ovid Online – www.ovid.
com/) 1971 to January 2007**

Searched 8 February 2007.
Retrieved 1630 hits.

1. Pregnancy.de.
2. Puerperium.de.
3. "Postnatal health".de.
4. Postnatal care.de.
5. pregnancy.ti,ab.
6. pregnant.ti,ab.
7. prenatal.ti,ab.
8. pre-natal.ti,ab.
9. postnatal.ti,ab.
10. postnatal.ti,ab.
11. postpartum.ti,ab.
12. post-partum.ti,ab.
13. puerperal.ti,ab.
14. new mother$.ti,ab.
15. pre-pregnancy.ti,ab.
16. prepregnancy.ti,ab.
17. ante-natal.ti,ab.
18. antenatal.ti,ab.
19. antepartum.ti,ab.
20. ante-partum.ti,ab.
21. or/1–20
22. Depression.de.
23. Postnatal depression.de.
24. pnd.ti,ab.
25. blues.ti,ab.
26. depress$.ti,ab.
27. melancholia.ti,ab.
28. (anxiety or anxious).ti,ab.
29. Anxiety.de.
30. ppd.ti,ab.
31. or/22–30
32. screen$.ti,ab.
33. diagnos$.ti,ab.
34. detect$.ti,ab.
35. predict$.ti,ab.
36. aware$.ti,ab.
37. identif$.ti,ab.
38. Screening.de.
39. Edinburgh Postnatal Depression Scale.de.
40. Postnatal depression diagnosis.de.
41. Beck Depression Inventory.de.
42. (edinburgh adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
43. EPDS.ti,ab.
44. (Postpartum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
45. (Post-partum adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
46. PDSS.ti,ab.
47. (Bromley adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
48. BPDS.ti,ab.
49. (General Health adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
50. GHQ.ti,ab.
51. (Beck adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
52. BDI.ti,ab.
53. BAI.ti,ab.
54. (State adj2 anxiety adj2 depression).ti,ab.
55. SAD.ti,ab.
56. (Hospital adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
57. HADS.ti,ab.
58. (Hamilton adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
59. HRSD.ti,ab.
60. (Zung adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
61. SDS.ti,ab.
62. Profile of mood states.ti,ab.
63. POMS.ti,ab.
64. (Centre adj2 Epidemiological studies adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
65. (CES-D or CESD).ti,ab.
66. Symptom Checklist-90-revised.ti,ab.
67. SCL-90-R.ti,ab.
68. (Brief symptom adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
69. BSI.ti,ab.
70. ((Inventory or Questionnaire or scale or index or checklist or interview) adj5 depressive symptomatology).ti,ab.
71. IDS.ti,ab.
72. (Montgomery-Asberg adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
73. MADRS.ti,ab.
74. (Depressive Adjective adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
75. DACL.ti,ab.
76. (Schedule adj2 affective disorders adj2 schizophrenia).ti,ab.
77. SADS.ti,ab.
78. (State-Trait anxiety adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.

79. STAI.ti,ab.
80. (Brisbane postnatal adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
81. (Post-partum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
82. (Postpartum depression predictors adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
83. ((Depress$ or anxiety) adj5 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab.
84. antenatal psychosocial health assessment. ti,ab.
85. alpha.ti,ab.
86. or/32–85
87. 21 and 31 and 86

CENTRAL and DARE (Database of Abstracts of Reviews of Effects) (Cochrane Library – CD-ROM) Issue 1 2007

Searched 9 February 2007.
Retrieved 326 hits from CENTRAL and 52 hits from DARE.

#1   MeSH descriptor Pregnancy this term only
#2   MeSH descriptor Prenatal Care this term only
#3   MeSH descriptor Postnatal Care this term only
#4   pregnancy in All Text
#5   pregnant in All Text
#6   prenatal in All Text
#7   pre next natal in All Text
#8   postnatal in All Text
#9   post next natal in All Text
#10   postpartum in All Text
#11   post next partum in All Text
#12   puerperal in All Text
#13   new next mother* in All Text
#14   pre next pregnancy in All Text
#15   prepregnancy in All Text
#16   ante next natal in All Text
#17   antenatal in All Text
#18   antepartum in All Text
#19   ante next partum in All Text
#20   (#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19)
#21   MeSH descriptor Depression this term only
#22   MeSH descriptor Depression, Postpartum this term only

#23   "pnd" in All Text
#24   blues in All Text
#25   depress* in All Text
#26   MeSH descriptor Depressive Disorder this term only
#27   melancholia in All Text
#28   (anxiety in All Text or anxious in All Text)
#29   MeSH descriptor anxiety this term only
#30   "ppd" in All Text
#31   (#21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30)
#32   screen* in All Text
#33   diagnos* in All Text
#34   detect* in All Text
#35   predict* in All Text
#36   aware* in All Text
#37   identif* in All Text
#38   MeSH descriptor DIAGNOSIS this term only
#39   (edinburgh in All Text near/5 Inventory in All Text)
#40   (edinburgh in All Text near/5 questionnaire in All Text)
#41   (edinburgh in All Text near/5 scale in All Text)
#42   (edinburgh in All Text near/5 index in All Text)
#43   (edinburgh in All Text near/5 checklist in All Text)
#44   (edinburgh in All Text near/5 interview in All Text)
#45   EPDS in All Text
#46   (postpartum in All Text near/5 inventory in All Text)
#47   (postpartum in All Text near/5 questionnaire in All Text)
#48   (postpartum in All Text near/5 scale in All Text)
#49   (postpartum in All Text near/5 index in All Text)
#50   (postpartum in All Text near/5 checklist in All Text)
#51   (postpartum in All Text near/5 interview in All Text)
#52   (post next partum in All Text near/5 inventory in All Text)
#53   (post next partum in All Text near/5 questionnaire in All Text)
#54   (post next partum in All Text near/5 scale in All Text)
#55   (post next partum in All Text near/5 index in All Text)
#56   (post next partum in All Text near/5 checklist in All Text)
#57   (post next partum in All Text near/5 interview in All Text)

#58 "PDSS" in All Text
#59 (Bromley in All Text near/5 inventory in All Text)
#60 (Bromley in All Text near/5 questionnaire in All Text)
#61 (Bromley in All Text near/5 scale in All Text)
#62 (Bromley in All Text near/5 index in All Text)
#63 (Bromley in All Text near/5 checklist in All Text)
#64 (Bromley in All Text near/5 interview in All Text)
#65 "BPDS" in All Text
#66 (general next health in All Text near/5 inventory in All Text)
#67 (general next health in All Text near/5 questionnaire in All Text)
#68 (general next health in All Text near/5 scale in All Text)
#69 (general next health in All Text near/5 index in All Text)
#70 (general next health in All Text near/5 checklist in All Text)
#71 (general next health in All Text near/5 interview in All Text)
#72 "GHQ" in All Text
#73 (Beck in All Text near/5 inventory in All Text)
#74 (Beck in All Text near/5 questionnaire in All Text)
#75 (Beck in All Text near/5 scale in All Text)
#76 (Beck in All Text near/5 index in All Text)
#77 (Beck in All Text near/5 checklist in All Text)
#78 (Beck in All Text near/5 interview in All Text)
#79 "BDI" in All Text
#80 "BAI" in All Text
#81 (state in All Text near/2 anxiety in All Text near/2 depression in Al Text)
#82 "SAD" in All Text
#83 (hospital in All Text near/5 inventory in All Text)
#84 (hospital in All Text near/5 questionnaire in All Text)
#85 (hospital in All Text near/5 scale in All Text)
#86 (hospital in All Text near/5 index in All Text)
#87 (hospital in All Text near/5 checklist in All Text)
#88 (hospital in All Text near/5 interview in All Text)
#89 "HADS" in All Text
#90 (hamilton in All Text near/5 inventory in All Text)
#91 (hamilton in All Text near/5 questionnaire in All Text)

#92 (hamilton in All Text near/5 scale in All Text)
#93 (hamilton in All Text near/5 index in All Text)
#94 (hamilton in All Text near/5 checklist in All Text)
#95 (hamilton in All Text near/5 interview in All Text)
#96 "HRSD" in All Text
#97 (zung in All Text near/5 inventory in All Text)
#98 (zung in All Text near/5 questionnaire in All Text)
#99 (zung in All Text near/5 scale in All Text)
#100 (zung in All Text near/5 index in All Text)
#101 (zung in All Text near/5 checklist in All Text)
#102 (zung in All Text near/5 interview in All Text)
#103 "SDS" in All Text
#104 (profile in All Text near/3 mood next states in All Text)
#105 "POMS" in All Text
#106 (centre in All Text near/5 epidemiological next studies in All Text near/5 inventory in All Text)
#107 (centre in All Text near/5 epidemiological next studies in All Text near/5 questionnaire in All Text)
#108 (centre in All Text near/5 epidemiological next studies in All Text near/5 scale in All Text)
#109 (centre in All Text near/5 epidemiological next studies in All Text near/5 index in All Text)
#110 (centre in All Text near/5 epidemiological next studies in All Text near/5 checklist in All Text)
#111 (centre in All Text near/5 epidemiological next studies in All Text near/5 interview in All Text)
#112 (CES next D in All Text or CESD in All Text)
#113 symptom next checklist next 90 next revised in All Text
#114 SCL next 90 next R in All Text
#115 (brief next symptom in All Text near/5 inventory in All Text)
#116 (brief next symptom in All Text near/5 questionnaire in All Text)
#117 (brief next symptom in All Text near/5 scale in All Text)
#118 (brief next symptom in All Text near/5 index in All Text)
#119 (brief next symptom in All Text near/5 checklist in All Text)
#120 (brief next symptom in All Text near/5 interview in All Text)

#121   "BSI" in All Text

#122   (interview in All Text near/5 depressive next symptomatology in All Text)

#123   (checklist in All Text near/5 depressive next symptomatology in All Text)

#124   (index in All Text near/5 depressive next symptomatology in All Text)

#125   (scale in All Text near/5 depressive next symptomatology in All Text)

#126   (questionnaire in All Text near/5 depressive next symptomatology in All Text)

#127   (inventory in All Text near/5 depressive next symptomatology in All Text)

#128   "IDS" in All Text

#129   (montgomery next asberg in All Text near/5 inventory in All Text)

#130   (montgomery next asberg in All Text near/5 questionnaire in All Text)

#131   (montgomery next asberg in All Text near/5 scale in All Text)

#132   (montgomery next asberg in All Text near/5 index in All Text)

#133   (montgomery next asberg in All Text near/5 checklist in All Text)

#134   (montgomery next asberg in All Text near/5 interview in All Text)

#135   "MADRS" in All Text

#136   (depressive in All Text and (adjective in All Text near/5 inventory in All Text))

#137   (depressive in All Text and (adjective in All Text near/5 questionnaire in All Text))

#138   (depressive in All Text and (adjective in All Text near/5 scale in All Text))

#139   (depressive in All Text and (adjective in All Text near/5 index in All Text))

#140   (depressive in All Text and (adjective in All Text near/5 checklist in All Text))

#141   (depressive in All Text and (adjective in All Text near/5 interview in All Text))

#142   "DACL" in All Text 9

#143   (schedule in All Text near/2 affective next disorders in All Text near/2 schizophrenia in All Text)

#144   "SADS" in All Text

#145   (state next trait next anxiety in All Text near/5 inventory in All Text)

#146   (state next trait next anxiety in All Text near/5 questionnaire in All Text)

#147   (state next trait next anxiety in All Text near/5 scale in All Text)

#148   (state next trait next anxiety in All Text near/5 index in All Text)

#149   (state next trait next anxiety in All Text near/5 checklist in All Text)

#150   (state next trait next anxiety in All Text near/5 interview in All Text)

#151   "STAI" in All Text

#152   (brisbane next postnatal in All Text near/5 inventory in All Text)

#153   (brisbane next postnatal in All Text near/5 questionnaire in All Text)

#154   (brisbane next postnatal in All Text near/5 scale in All Text)

#155   (brisbane next postnatal in All Text near/5 index in All Text)

#156   (brisbane next postnatal in All Text near/5 checklist in All Text)

#157   (brisbane next postnatal in All Text near/5 interview in All Text)

#158   ((post next partum next depression next predictors in All Text near/6/5 in All Text) and inventory in All Text)

#159   (post next partum next depression next predictors in All Text near/6 questionnaire in All Text)

#160   (post next partum next depression next predictors in All Text near/6 scale in All Text)

#161   (post next partum next depression next predictors in All Text near/6 index in All Text)

#162   (post next partum next depression next predictors in All Text near/6 checklist in All Text)

#163   (post next partum next depression next predictors in All Text near/6 interview in All Text)

#164   (postpartum next depression next predictors in All Text near/5 inventory in All Text)

#165   (postpartum next depression next predictors in All Text near/5 questionnaire in All Text)

#166   (postpartum next depression next predictors in All Text near/5 scale in All Text)

#167   (postpartum next depression next predictors in All Text near/5 index in All Text)

#168   (postpartum next depression next predictors in All Text near/5 checklist in All Text)

#169   (postpartum next depression next predictors in All Text near/5 interview in All Text)

#170   (depress* in All Text near/5 inventory in All Text)

#171   (depress* in All Text near/5 questionnaire in All Text)

#172   (depress* in All Text near/5 scale in All Text)

#173   (depress* in All Text near/5 index in All Text)

#174   (depress* in All Text near/5 checklist in All Text)

#175   (depress* in All Text near/5 interview in All Text)

#176   (anxiety in All Text near/5 inventory in All Text)

#177 (anxiety in All Text near/5 questionnaire in All Text)

#178 (anxiety in All Text near/5 scale in All Text)

#179 (anxiety in All Text near/5 index in All Text)

#180 (anxiety in All Text near/5 checklist in All Text)

#181 (anxiety in All Text near/5 interview in All Text)

#182 MeSH descriptor QUESTIONNAIRES this term only

#183 MeSH descriptor INTERVIEWS this term only

#184 antenatal next psychosocial next health next assessment in All Text

#185 alpha in All Text

#186 (#32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65 or #66 or #67 or #68 or #69 or #70 or #71 or #72 or #73 or #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82 or #83 or #84 or #85 or #86 or #87 or #88 or #89 or #90 or #91 or #92 or #93 or #94 or #95 or #96 or #97 or #98 or #99 or #100 or #101 or #102 or #103 or #104 or #105 or #116 or #117 or #118 or #119 or #120 or #121 or #122 or #123 or #124 or #125 or #126 or #127 or #128 or #129 or #130 or #131 or #132 or #133 or #134 or #135 or #136 or #137 or #138 or #139 or #140 or #141 or #142 or #143 or #144 or #145 or #146 or #147 or #148 or #149 or #150 or #151 or #152 or #153 or #154 or #155 or #156 or #157 or #158 or #159 or #160 or #161 or #162 or #163 or #164 or #165 or #166 or #167 or #168 or #169 or #170 or #171 or #172 or #173 or #174 or #175 or #176 or #177 or #178 or #179 or #180 or #181 or #182 or #183 or #184 or #185)

#187 (#20 and #31 and #186)

CDSR (Cochrane Database of Systematic Reviews) (Cochrane Library – CD-ROM) Issue 1 2007

Searched 9 February 2007.
Retrieved 12 hits.

#188 MeSH descriptor Pregnancy this term only

#189 MeSH descriptor Prenatal Care this term only

#190 MeSH descriptor Postnatal Care this term only

#191 (pregnancy in Record Title or pregnancy in Abstract)

#192 (pregnant in Record Title or pregnant in Abstract)

#193 (prenatal in Record Title or prenatal in Abstract)

#194 (pre next natal in Record Title or pre next natal in Abstract)

#195 (postnatal in Record Title or postnatal in Abstract)

#196 (post next natal in Record Title or post next natal in Abstract)

#197 (postpartum in Record Title or postpartum in Abstract)

#198 (post next partum in Record Title or post next partum in Abstract)

#199 (puerperal in Record Title or puerperal in Abstract)

#200 (new next mother* in Record Title or new next mother* in Abstract)

#201 (pre next pregnancy in Record Title or pre next pregnancy in Abstract)

#202 (prepregnancy in Record Title or prepregnancy in Abstract)

#203 (ante next natal in Record Title or ante next natal in Abstract) 10

#204 (antenatal in Record Title or antenatal in Abstract)

#205 (antepartum in Record Title or antepartum in Abstract)

#206 (ante next partum in Record Title or ante next partum in Abstract)

#207 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19)

#208 MeSH descriptor Depression this term only

#209 MeSH descriptor Depression, Postpartum this term only

#210 "pnd" in Record Title or "pnd" in Abstract)

#211 (blues in Record Title or blues in Abstract)

#212 (depress* in Record Title or depress* in Abstract)

#213 MeSH descriptor Depressive Disorder this term only

#214 (melancholia in Record Title or melancholia in Abstract)

#215 ((anxiety in Record Title or anxious in Record Title) or (anxiety in Abstract or anxious in Abstract))

#216 MeSH descriptor anxiety this term only

#217 ("ppd" in Record Title or "ppd" in Abstract)

#218 (#21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30)

#219 (screen* in Record Title or screen* in Abstract)

#220 (diagnos* in Record Title or diagnos* in Abstract)

#221 (detect* in Record Title or detect* in Abstract)

#222 (predict* in Record Title or predict* in Abstract)

#223 (aware* in Record Title or aware* in Abstract)

#224 (identif* in Record Title or identif* in Abstract)

#225 MeSH descriptor DIAGNOSIS this term only

#226 ((edinburgh in Record Title near/5 Inventory in Record Title) or (edinburgh in Abstract near/5 Inventory in Abstract))

#227 ((edinburgh in Record Title near/5 questionnaire in Record Title) or (edinburgh in Abstract near/5 questionnaire in Abstract))

#228 ((edinburgh in Record Title near/5 scale in Record Title) or (edinburgh in Abstract near/5 scale in Abstract))

#229 ((edinburgh in Record Title near/5 index in Record Title) or (edinburgh in Abstract near/5 index in Abstract))

#230 ((edinburgh in Record Title near/5 checklist in Record Title) or (edinburgh in Abstract near/5 checklist in Abstract))

#231 ((edinburgh in Record Title near/5 interview in Record Title) or (edinburgh in Abstract near/5 interview in Abstract))

#232 (EPDS in Record Title or EPDS in Abstract)

#233 ((postpartum in Record Title near/5 inventory in Record Title) or (postpartum in Abstract near/5 inventory in Abstract))

#234 ((postpartum in Record Title near/5 questionnaire in Record Title) or (postpartum in Abstract near/5 questionnaire in Abstract))

#235 ((postpartum in Record Title near/5 scale in Record Title) or (postpartum in Abstract near/5 scale in Abstract))

#236 ((postpartum in Record Title near/5 index in Record Title) or (postpartum in Abstract near/5 index in Abstract))

#237 ((postpartum in Record Title near/5 checklist in Record Title) or (postpartum in Abstract near/5 checklist in Abstract))

#238 ((postpartum in Record Title near/5 interview in Record Title) or (postpartum in Abstract near/5 interview in Abstract))

#239 ((post next partum in Record Title near/5 inventory in Record Title) or (post next partum in Abstract near/5 inventory in Abstract))

#240 ((post next partum in Record Title near/5 questionnaire in Record Title) or (post next partum in Abstract near/5 questionnaire in Abstract))

#241 ((post next partum in Record Title near/5 scale in Record Title) or (post next partum in Abstract near/5 scale in Abstract))

#242 ((post next partum in Record Title near/5 index in Record Title) or (post next partum in Abstract near/5 index in Abstract))

#243 ((post next partum in Record Title near/5 checklist in Record Title) or (post next partum in Abstract near/5 checklist in Abstract))

#244 ((post next partum in Record Title near/5 interview in Record Title) or (post next partum in Abstract near/5 interview in Abstract))

#245 ("PDSS" in Record Title or "PDSS" in Abstract)

#246 ((Bromley in Record Title near/5 inventory in Record Title) or (Bromley in Abstract near/5 inventory in Abstract))

#247 ((Bromley in Record Title near/5 questionnaire in Record Title) or (Bromley in Abstract near/5 questionnaire in Abstract))

#248 ((Bromley in Record Title near/5 scale in Record Title) or (Bromley in Abstract near/5 scale in Abstract))

#249 ((Bromley in Record Title near/5 index in Record Title) or (Bromley in Abstract near/5 index in Abstract))

#250 ((Bromley in Record Title near/5 checklist in Record Title) or (Bromley in Abstract near/5 checklist in Abstract))

#251 ((Bromley in Record Title near/5 interview in Record Title) or (Bromley in Abstract near/5 interview in Abstract))

#252 ("BPDS" in Record Title or "BPDS" in Abstract)

#253 ((general next health in Record Title near/5 inventory in Record Title) or (general next health in Abstract near/5 inventory in Abstract))

#254 ((general next health in Record Title near/5 questionnaire in Record Title) or (general next health in Abstract near/5 questionnaire in Abstract))

#255 ((general next health in Record Title near/5 scale in Record Title) or (general next health in Abstract near/5 scale in Abstract))

#256 ((general next health in Record Title near/5 scale in Record Title) or (general next health in Abstract near/5 scale in Abstract))

#257 ((general next health in Record Title near/5 checklist in Record Title) or (general

next health in Abstract near/5 checklist in Abstract))

#258 ((general next health in Record Title near/5 interview in Record Title) or (general next health in Abstract near/5 interview in Abstract))

#259 ("GHQ" in Record Title or "GHQ" in Abstract)

#260 ((Beck in Record Title near/5 inventory in Record Title) or (Beck in Abstract near/5 inventory in Abstract))

#261 ((Beck in Record Title near/5 questionnaire in Record Title) or (Beck in Abstract near/5 questionnaire in Abstract))

#262 ((Beck in Record Title near/5 scale in Record Title) or (Beck in Abstract near/5 scale in Abstract))

#263 ((Beck in Record Title near/5 index in Record Title) or (Beck in Abstract near/5 index in Abstract))

#264 ((Beck in Record Title near/5 checklist in Record Title) or (Beck in Abstract near/5 checklist in Abstract))

#265 ((Beck in Record Title near/5 interview in Record Title) or (Beck in Abstract near/5 interview in Abstract))

#266 ("BDI" in Record Title or "BDI" in Abstract)

#267 ("BAI" in Record Title or "BAI" in Abstract)

#268 ((state in Record Title near/2 anxiety in Record Title near/2 depression in Record Title) or (state in Abstract near/2 anxiety in Abstract near/2 depression in Abstract))

#269 ("SAD" in Record Title or "SAD" in Abstract)

#270 ((hospital in Record Title near/5 inventory in Record Title) or (hospital in Abstract near/5 inventory in Abstract))

#271 ((hospital in Record Title near/5 questionnaire in Record Title) or (hospital in Abstract near/5 questionnaire in Abstract))

#272 ((hospital in Record Title near/5 scale in Record Title) or (hospital in Abstract near/5 scale in Abstract))

#273 ((hospital in Record Title near/5 index in Record Title) or (hospital in Abstract near/5 index in Abstract))

#274 ((hospital in Record Title near/5 checklist in Record Title) or (hospital in Abstract near/5 checklist in Abstract))

#275 ((hospital in Record Title near/5 interview in Record Title) or (hospital in Abstract near/5 interview in Abstract))

#276 ("HADS" in Record Title or "HADS" in Abstract)

#277 ((hamilton in Record Title near/5 inventory in Record Title) or (hamilton in Abstract near/5 inventory in Abstract))

#278 ((hamilton in Record Title near/5 questionnaire in Record Title) or (hamilton in Abstract near/5 questionnaire in Abstract))

#279 ((hamilton in Record Title near/5 scale in Record Title) or (hamilton in Abstract near/5 scale in Abstract))

#280 ((hamilton in Record Title near/5 index in Record Title) or (hamilton in Abstract near/5 index in Abstract))

#281 ((hamilton in Record Title near/5 checklist in Record Title) or (hamilton in Abstract near/5 checklist in Abstract))

#282 ((hamilton in Record Title near/5 interview in Record Title) or (hamilton in Abstract near/5 interview in Abstract))

#283 ("HRSD" in Record Title or "HRSD" in Abstract)

#284 ((zung in Record Title near/5 inventory in Record Title) or (zung in Abstract near/5 inventory in Abstract))

#285 ((zung in Record Title near/5 questionnaire in Record Title) or (zung in Abstract near/5 questionnaire in Abstract))

#286 ((zung in Record Title near/5 scale in Record Title) or (zung in Abstract near/5 scale in Abstract))

#287 ((zung in Record Title near/5 index in Record Title) or (zung in Abstract near/5 index in Abstract))

#288 ((zung in Record Title near/5 checklist in Record Title) or (zung in Abstract near/5 checklist in Abstract))

#289 ((zung in Record Title near/5 interview in Record Title) or (zung in Abstract near/5 interview in Abstract))

#290 ("SDS" in Record Title or "SDS" in Abstract)

#291 ((profile in Record Title near/3 mood next states in Record Title) or (profile in Abstract near/3 mood next states in Abstract))

#292 ("POMS" in Record Title or "POMS" in Abstract)

#293 ((centre in Record Title near/5 epidemiological next studies in Record Title near/5 inventory in Record Title) or (centre in Abstract near/5 epidemiological next studies in Abstract near/5 inventory in Abstract))

#294 ((centre in Record Title near/5 epidemiological next studies in Record Title near/5 questionnaire in Record Title) or (centre in Abstract near/5 epidemiological next studies in Abstract near/5 questionnaire in Abstract))

#295 ((centre in Record Title near/5 epidemiological next studies in Record Title near/5 scale in Record Title) or (centre in

Abstract near/5 epidemiological next studies in Abstract near/5 scale in Abstract))

#296 ((centre in Record Title near/5 epidemiological next studies in Record Title near/5 index in Record Title) or (centre in Abstract near/5 epidemiological next studies in Abstract near/5 index in Abstract))

#297 ((centre in Record Title near/5 epidemiological next studies in Record Title near/5 checklist in Record Title) or (centre in Abstract near/5 epidemiological next studies in Abstract near/5 checklist in Abstract))

#298 ((centre in Record Title near/5 epidemiological next studies in Record Title near/5 interview in Record Title) or (centre in Abstract near/5 epidemiological next studies in Abstract near/5 interview in Abstract))

#299 ((CES next D in Record Title or CESD in Record Title) or (CES next D in Abstract or CESD in Abstract)) 75 edit delete

#300 (symptom next checklist next 90 next revised in Record Title or symptom next checklist next 90 next revised in Abstract)

#301 (SCL next 90 next R in Record Title or SCL next 90 next R in Abstract)

#302 ((brief next symptom in Record Title near/5 inventory in Record Title) or (brief next symptom in Abstract near/5 inventory in Abstract))

#303 ((brief next symptom in Record Title near/5 questionnaire in Record Title) or (brief next symptom in Abstract near/5 questionnaire in Abstract))

#304 ((brief next symptom in Record Title near/5 scale in Record Title) or (brief next symptom in Abstract near/5 scale in Abstract))

#305 ((brief next symptom in Record Title near/5 index in Record Title) or (brief next symptom in Abstract near/5 index in Abstract))

#306 ((brief next symptom in Record Title near/5 checklist in Record Title) or (brief next symptom in Abstract near/5 checklist in Abstract))

#307 ((brief next symptom in Record Title near/5 interview in Record Title) or (brief next symptom in Abstract near/5 interview in Abstract))

#308 ("BSI" in Record Title or "BSI" in Abstract)

#309 ((interview in Record Title near/5 depressive next symptomatology in Record Title) or (interview in Abstract near/5 depressive next symptomatology in Abstract))

#310 ((checklist in Record Title near/5 depressive next symptomatology in Record Title) or

(checklist in Abstract near/5 depressive next symptomatology in Abstract))

#311 ((index in Record Title near/5 depressive next symptomatology in Record Title) or (index in Abstract near/5 depressive next symptomatology in Abstract))

#312 ((scale in Record Title near/5 depressive next symptomatology in Record Title) or (scale in Abstract near/5 depressive next symptomatology in Abstract))

#313 ((questionnaire in Record Title near/5 depressive next symptomatology in Record Title) or (questionnaire in Abstract near/5 depressive next symptomatology in Abstract))

#314 ((inventory in Record Title near/5 depressive next symptomatology in Record Title) or (inventory in Abstract near/5 depressive next symptomatology in Abstract))

#315 ("IDS" in Record Title or "IDS" in Abstract)

#316 ((montgomery next asberg in Record Title near/5 inventory in Record Title) or (montgomery next asberg in Abstract near/5 inventory in Abstract))

#317 ((montgomery next asberg in Record Title near/5 questionnaire in Record Title) or (montgomery next asberg in Abstract near/5 questionnaire in Abstract))

#318 ((montgomery next asberg in Record Title near/5 scale in Record Title) or (montgomery next asberg in Abstract near/5 scale in Abstract))

#319 ((montgomery next asberg in Record Title near/5 index in Record Title) or (montgomery next asberg in Abstract near/5 index in Abstract))

#320 ((montgomery next asberg in Record Title near/5 checklist in Record Title) or (montgomery next asberg in Abstract near/5 checklist in Abstract))

#321 ((montgomery next asberg in Record Title near/5 interview in Record Title) or (montgomery next asberg in Abstract near/5 interview in Abstract))

#322 ("MADRS" in Record Title or "MADRS" in Abstract)

#323 ((depressive in Record Title and (adjective in Record Title near/5 inventory in Record Title)) or (depressive in Abstract and (adjective in Abstract near/5 inventory in Abstract)))

#324 ((depressive in Record Title and (adjective in Record Title near/5 questionnaire in Record Title)) or (depressive in Abstract and (adjective in Abstract near/5 questionnaire in Abstract)))

#325 ((depressive in Record Title and (adjective in Record Title near/5 scale in Record Title)) or (depressive in Abstract and (adjective in Abstract near/5 scale in Abstract)))

#326 ((depressive in Record Title and (adjective in Record Title near/5 index in Record Title)) or (depressive in Abstract and (adjective in Abstract near/5 index in Abstract)))

#327 ((depressive in Record Title and (adjective in Record Title near/5 checklist in Record Title)) or (depressive in Abstract and (adjective in Abstract near/5 checklist in Abstract)))

#328 ((depressive in Record Title and (adjective in Record Title near/5 interview in Record Title)) or (depressive in Abstract and (adjective in Abstract near/5 interview in Abstract)))

#329 ("DACL" in Record Title or "DACL" in Abstract)

#330 ((schedule in Record Title near/2 affective next disorders in Record Title near/2 schizophrenia in Record Title) or (schedule in Abstract near/2 affective next disorders in Abstract near/2 schizophrenia in Abstract))

#331 ("SADS" in Record Title or "SADS" in Abstract)

#332 ((state next trait next anxiety in Record Title near/5 inventory in Record Title) or (state next trait next anxiety in Abstract near/5 inventory in Abstract))

#333 ((state next trait next anxiety in Record Title near/5 questionnaire in Record Title) or (state next trait next anxiety in Abstract near/5 questionnaire in Abstract))

#334 ((state next trait next anxiety in Record Title near/5 scale in Record Title) or (state next trait next anxiety in Abstract near/5 scale in Abstract))

#335 ((state next trait next anxiety in Record Title near/5 index in Record Title) or (state next trait next anxiety in Abstract near/5 index in Abstract))

#336 ((state next trait next anxiety in Record Title near/5 checklist in Record Title) or (state next trait next anxiety in Abstract near/5 checklist in Abstract))

#337 ((state next trait next anxiety in Record Title near/5 interview in Record Title) or (state next trait next anxiety in Abstract near/5 interview in Abstract))

#338 ("STAI" in Record Title or "STAI" in Abstract)

#339 (brisbane next postnatal in All Text near/5 inventory in All Text)

#340 (brisbane next postnatal in All Text near/5 questionnaire in All Text)

#341 (brisbane next postnatal in All Text near/5 scale in All Text)

#342 (brisbane next postnatal in All Text near/5 index in All Text)

#343 (brisbane next postnatal in All Text near/5 checklist in All Text)

#344 (brisbane next postnatal in All Text near/5 interview in All Text)

#345 ((post next partum next depression next predictors in All Text near/6/5 in All Text) and inventory in All Text)

#346 (post next partum next depression next predictors in All Text near/6 questionnaire in All Text)

#347 (post next partum next depression next predictors in All Text near/6 scale in All Text)

#348 (post next partum next depression next predictors in All Text near/6 index in All Text)

#349 (post next partum next depression next predictors in All Text near/6 checklist in All Text)

#350 (post next partum next depression next predictors in All Text near/6 interview in All Text)

#351 (postpartum next depression next predictors in All Text near/5 inventory in All Text)

#352 (postpartum next depression next predictors in All Text near/5 questionnaire in All Text)

#353 (postpartum next depression next predictors in All Text near/5 scale in All Text)

#354 (postpartum next depression next predictors in All Text near/5 index in All Text)

#355 (postpartum next depression next predictors in All Text near/5 checklist in All Text)

#356 (postpartum next depression next predictors in All Text near/5 interview in All Text)

#357 ((depress* in Record Title near/5 inventory in Record Title) or (depress* in Abstract near/5 inventory in Abstract))

#358 ((depress* in Record Title near/5 questionnaire in Record Title) or (depress* in Abstract near/5 questionnaire in Abstract))

#359 ((depress* in Record Title near/5 scale in Record Title) or (depress* in Abstract near/5 scale in Abstract))

#360 ((depress* in Record Title near/5 index in Record Title) or (depress* in Abstract near/5 index in Abstract))

#361 ((depress* in Record Title near/5 checklist in Record Title) or (depress* in Abstract near/5 checklist in Abstract))

#362 ((depress* in Record Title near/5 interview in Record Title) or (depress* in Abstract near/5 interview in Abstract))

#363 ((anxiety in Record Title near/5 inventory in Record Title) or (anxiety in Abstract near/5 inventory in Abstract))

#364 ((anxiety in Record Title near/5 questionnaire in Record Title) or (anxiety in Abstract near/5 questionnaire in Abstract))

#365 ((anxiety in Record Title near/5 scale in Record Title) or (anxiety in Abstract near/5 scale in Abstract))

#366 ((anxiety in Record Title near/5 index in Record Title) or (anxiety in Abstract near/5 index in Abstract))

#367 ((anxiety in Record Title near/5 checklist in Record Title) or anxiety in Abstract near/5 checklist in Abstract))

#368 (anxiety in All Text near/5 interview in All Text)

#369 MeSH descriptor QUESTIONNAIRES this term only

#370 MeSH descriptor INTERVIEWS this term only

#371 antenatal next psychosocial next health next assessment in All Text

#372 (alpha in Record Title or alpha in Abstract)

#373 (#32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65 or #66 or #67 or #68 or #69 or #70 or #71 or #72 or #73 or #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82 or #83 or #84 or #85 or #86 or #87 or #88 or #89 or #90 or #91 or #92 or #93 or #94 or #95 or #96 or #97 or #98 or #99 or #100 or #101 or #102 or #103 or #104 or #105 or #116 or #117 or #118 or #119 or #120 or #121 or #122 or #123 or #124 or #125 or #126 or #127 or #128 or #129 or #130 or #131 or #132 or #133 or #134 or #135 or #136 or #137 or #138 or #139 or #140 or #141 or #142 or #143 or #144 or #145 or #146 or #147 or #148 or #149 or #150 or #151 or #152 or #153 or #154 or #155 or #156 or #157 or #158 or #159 or #160 or #161 or #162 or #163 or #164 or #165 or #166 or #167 or #168 or #169 or #170 or #171 or #172 or #173 or #174 or #175 or #176 or #177 or #178 or #179 or #180 or #181 or #182 or #183 or #184 or #185)

#374 (#20 and #31 and #186)

Science Citation Index (SSCI) (Web Of Knowledge – http://wos.mimas.ac.uk/) 1900 to present day

Searched 12 February 2007.

Retrieved 2870 hits.

All lines limited as follows: DocType = All document types; Language = All languages; Database = SCI-EXPANDED; Timespan = 1900–2007:

#1 TS=(pregnancy)
#2 TS=(pregnant)
#3 TS=(prenatal)
#4 TS=(pre-natal)
#5 TS=(postnatal)
#6 TS=(postnatal)
#7 TS=(postpartum)
#8 TS=(post-partum)
#9 TS=("new mother*")
#10 TS=(pre-pregnancy)
#11 TS=(prepregnancy)
#12 TS=(ante-natal)
#13 TS=(antenatal)
#14 TS=(antepartum)
#15 TS=(ante-partum)
#16 TS=(puerperal)
#17 #16 OR #15 OR #14 OR #13 OR #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1
#18 TS=(pnd)
#19 TS=(blues)
#20 TS=(depress*)
#21 TS=(melancholia)
#22 TS=(anxiety or anxious)
#23 TS=(ppd)
#24 #23 OR #22 OR #21 OR #20 OR #19 OR #18
#25 TS=(screen*)
#26 TS=(diagnos*)
#27 TS=(detect*)
#28 TS=(predict*)
#29 TS=(aware*)
#30 TS=(identif*)
#31 TS=(edinburgh same (inventory or questionnaire or scale or index or checklist or interview))
#32 TS=(EPDS)
#33 TS=(Postpartum same (Inventory or Questionnaire or scale or index or checklist or interview))

#34 TS=(Post-partum same (Inventory or Questionnaire or scale or index or checklist or interview))
#35 TS=(PDSS)
#36 TS=(Bromley same (Inventory or Questionnaire or scale or index or checklist or interview))
#37 TS=(BPDS)
#38 TS=("General Health" same (Inventory or Questionnaire or scale or index or checklist or interview))
#39 TS=(GHQ)
#40 TS=(Beck same (Inventory or Questionnaire or scale or index or checklist or interview))
#41 TS=(BDI)
#42 TS=(BAI)
#43 TS=(State same anxiety same depression)
#44 TS=(SAD)
#45 TS=(Hospital same (Inventory or Questionnaire or scale or index or checklist or interview))
#46 TS=(HADS)
#47 TS=(Hamilton same (Inventory or Questionnaire or scale or index or checklist or interview))
#48 TS=(HRSD)
#49 TS=(Zung same (Inventory or Questionnaire or scale or index or checklist or interview))
#50 TS=(SDS)
#51 TS=("Profile of mood states")
#52 TS=(POMS)
#53 TS=(Centre same "Epidemiological studies" same (Inventory or Questionnaire or scale or index or checklist or interview))
#54 TS=(CES-D or CESD)
#55 TS=("Symptom Checklist-90-revised")
#56 TS=(SCL-90-R)
#57 TS=("Brief symptom" same (Inventory or Questionnaire or scale or index or checklist or interview))
#58 TS=(BSI)
#59 TS=((Inventory or Questionnaire or scale or index or checklist or interview) same "depressive symptomatology")
#60 TS=(IDS)
#61 TS=(Montgomery-Asberg same (Inventory or Questionnaire or scale or index or checklist or interview))
#62 TS=(MADRS)
#63 TS=("Depressive Adjective" same (Inventory or Questionnaire or scale or index or checklist or interview))
#64 TS=(DACL)
#65 TS=(Schedule same affective disorders same schizophrenia)

#66 TS=(SADS)
#67 TS=("State-Trait anxiety" same (Inventory or Questionnaire or scale or index or checklist or interview))
#68 TS=(STAI)
#69 TS=("Brisbane postnatal" same (Inventory or Questionnaire or scale or index or checklist or interview))
#70 TS=("Post-partum depression predictors" same (Inventory or Questionnaire or scale or index or checklist or interview))
#71 TS=("Postpartum depression predictors" same (Inventory or Questionnaire or scale or index or checklist or interview))
#72 TS=((Depress* or anxiety) same (Inventory or Questionnaire or scale or index or checklist or interview))
#73 TS=("antenatal psychosocial health assessment")
#74 TS=(alpha)
#75 #74 OR #73 OR #72 OR #71 OR #70 OR #69 OR #68 OR #67 OR #66 OR #65 OR #64 OR #63 OR #62 OR #61 OR #60 OR #59 OR #58 OR #57 OR #56 OR #55 OR #54 OR #53 OR #52 OR #51 OR #50 OR #49 OR #48 OR #47 OR #46 OR #45 OR #44 OR #43 OR #42 OR #41 OR #40 OR #39 OR #38 OR #37 OR #36 OR #35 OR #34 OR #33 OR #32 OR #31 OR #30 OR #29 OR #28 OR #27 OR #26 OR #25
#76 #75 AND #24 AND #17

NRR (www.nrr.nhs.uk/) 2007 Issue 1

Searched 12 February 2007.
Retrieved 377 hits.

#1.   PRENATAL CARE single term (MeSH)
#2.   PREGNANCY check tag (MeSH)
#3.   POSTNATAL CARE single term (MeSH)
#4.   pregnancy
#5.   pregnant
#6.   prenatal
#7.   (pre next natal)
#8.   postnatal
#9.   (post next natal)
#10. postpartum
#11. (post next partum)
#12. puerperal
#13. (new next mother*)
#14. (pre next pregnancy)
#15. prepregnancy
#16. (ante next natal)
#17. antenatal
#18. antepartum

#19. (ante next partum)
#20. (#1 or #2 or #3 or #4 or #5 or #6 or #7
      or #8 or #9 or #10 or #11 or #12 or #13
      or #14 or #15 or #16 or #17 or #18 or
      #19)
#21. DEPRESSION single term (MeSH)
#22. DEPRESSION POSTPARTUM single term
      (MeSH)
#23. pnd
#24. blues
#25. depress*
#26. DEPRESSIVE DISORDER single term
      (MeSH)
#27. melancholia
#28. (anxiety or anxious)
#29. ANXIETY single term (MeSH)
#30. ppd
#31. (#21 or #22 or #23 or #24 or #25 or #26
      or #27 or #28 or #29 or #30)
#32. screen*
#33. diagnos*
#34. detect*
#35. predict*
#36. aware*
#37. identif*
#38. DIAGNOSIS single term (MeSH)
#39. (edinburgh near inventory)
#40. (edinburgh near questionnaire)
#41. (edinburgh near scale)
#42. (edinburgh near index)
#43. (edinburgh near checklist)
#44. (edinburgh near interview)
#45. epds
#46. (postpartum near inventory)
#47. (postpartum near questionnaire)
#48. (postpartum near scale)
#49. (postpartum near index)
#50. (postpartum near checklist)
#51. (postpartum near interview)
#52. ((post next partum) near inventory)
#53. ((post next partum) near questionnaire)
#54. ((post next partum) near scale)
#55. ((post next partum) near index)
#56. ((post next partum) near checklist)
#57. ((post next partum) near interview)
#58. pdss
#59. (bromley near inventory)
#60. (bromley near questionnaire)
#61. (bromley near scale)
#62. (bromley near index)
#63. (bromley near checklist)
#64. (bromley near interview)
#65. bpds
#66. ((general next health) near inventory)
#67. ((general next health) near questionnaire)
#68. ((general next health) near scale)

#69. ((general next health) near index)
#70. ((general next health) near checklist)
#71. ((general next health) near interview)
#72. ghq
#73. (beck near inventory)
#74. (beck near questionnaire)
#75. (beck near scale)
#76. (beck near index)
#77. (beck near checklist)
#78. (beck near interview)
#79. bdi
#80. bai
#81. (state near anxiety near depression)
#82. sad
#83. (hospital near inventory)
#84. (hospital near questionnaire)
#85. (hospital near scale)
#86. (hospital near index)
#87. (hospital near checklist)
#88. (hospital near interview)
#89. hads
#90. (hamilton near inventory)
#91. (hamilton near questionnaire)
#92. (hamilton near scale)
#93. (hamilton near index)
#94. (hamilton near checklist)
#95. (hamilton near interview)
#96. hrsd
#97. (zung near inventory)
#98. (zung near questionnaire)
#99. (zung near scale)
#100. (zung near index)
#101. (zung near checklist)
#102. (zung near interview)
#103. sds
#104. (profile near (mood next states))
#105. poms
#106. (centre near (epidemiological next studies)
      near inventory)
#107. (centre near (epidemiological next studies)
      near questionnaire)
#108. (centre near (epidemiological next studies)
      near scale)
#109. (centre near (epidemiological next studies)
      near index)
#110. (centre near (epidemiological next studies)
      near checklist)
#111. (centre near (epidemiological next studies)
      near interview)
#112. ((ces next d) or cesd)
#113. (symptom next checklist next 90 next
      revised)
#114. (scl next 90)
#115. ((brief next symptom) near inventory)
#116. ((brief next symptom) near questionnaire)
#117. ((brief next symptom) near scale)

#118. ((brief next symptom) near index)
#119. ((brief next symptom) near checklist)
#120. ((brief next symptom) near interview)
#121. bsi
#122. (interview near (depressive next symptomatology))
#123. (checklist near (depressive next symptomatology))
#124. (index near (depressive next symptomatology))
#125. (scale near (depressive next symptomatology))
#126. (questionnaire near (depressive next symptomatology))
#127. (inventory near (depressive next symptomatology))
#128. ids
#129. ((montgomery next asberg) near inventory)
#130. ((montgomery next asberg) near questionnaire)
#131. ((montgomery next asberg) near scale)
#132. ((montgomery next asberg) near index)
#133. ((montgomery next asberg) near checklist)
#134. ((montgomery next asberg) near interview)
#135. madrs
#136. (depressive near adjective near inventory)
#137. (depressive near adjective near questionnaire)
#138. (depressive near adjective near scale)
#139. (depressive near adjective near index)
#140. (depressive near adjective near checklist)
#141. (depressive near adjective near interview)
#142. dacl
#143. (schedule near (affective next disorders) near schizophrenia)
#144. sads
#145. ((state next trait next anxiety) near inventory)
#146. ((state next trait next anxiety) near questionnaire)
#147. ((state next trait next anxiety) near scale)
#148. ((state next trait next anxiety) near index)
#149. ((state next trait next anxiety) near index)
#150. ((state next trait next anxiety) near interview)
#151. stai
#152. ((brisbane next postnatal) near inventory)
#153. ((brisbane next postnatal) near questionnaire)
#154. ((brisbane next postnatal) near scale)
#155. ((brisbane next postnatal) near index)
#156. ((brisbane next postnatal) near checklist)
#157. ((brisbane next postnatal) near interview)
#158. ((post next partum next depression next predictors) near inventory)

#159. ((post next partum next depression next predictors) near questionnaire)
#160. ((post next partum next depression next predictors) near scale)
#161. ((post next partum next depression next predictors) near index)
#162. ((post next partum next depression next predictors) near checklist)
#163. ((post next partum next depression next predictors) near interview)
#164. ((postpartum next depression next predictors) near inventory)
#165. ((postpartum next depression next predictors) near questionnaire)
#166. ((postpartum next depression next predictors) near scale)
#167. ((postpartum next depression next predictors) near index)
#168. ((postpartum next depression next predictors) near checklist)
#169. ((postpartum next depression next predictors) near interview)
#170. (depress* near inventory)
#171. (depress* near questionnaire)
#172. (depress* near scale)
#173. (depress* near index)
#174. (depress* near checklist)
#175. (depress* near interview)
#176. (anxiety near inventory)
#177. (anxiety near questionnaire)
#178. (anxiety near scale)
#179. (anxiety near index)
#180. (anxiety near checklist)
#181. interview
#182. QUESTIONNAIRES single term (MeSH)
#183. INTERVIEWS single term (MeSH)
#184. (antenatal next psychosocial next health next assessment)
#185. alpha
#186. (#32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65 or #66 or #67 or #68 or #69 or #70 or #71 or #72 or #73 or #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82 or #83 or #84 or #85 or #86 or #87 or #88 or #89 or #90 or #91 or #92 or #93 or #94 or #95 or #96 or #97 or #98 or #99 or #100 or #101 or #102 or #103 or #104 or #105 or #116 or #117 or #118 or #119 or #120 or #121 or #122 or #123 or #124 or #125 or #126 or #127 or #128 or #129 or #130 or #131 or #132 or

#133 or #134 or #135 or #136 or #137 or
#138 or #139 or #140 or #141 or #142 or
#143 or #144 or #145 or #146 or #147 or
#148 or #149 or #150 or #151 or #152 or
#153 or #154 or #155 or #156 or #157 or
#158 or #159 or #160 or #161 or #162 or
#163 or #164 or #165 or #166 or #167 or
#168 or #169 or #170 or #171 or #172 or
#173 or #174 or #175 or #176 or #177 or
#178 or #179 or #180 or #181 or #182 or
#183 or #184 or #185)

#187. (#20 and #31 and #186)

ReFeR (Research Findings Register)

Searched 12 February 2007.
Retrieved 26 hits.

The search interface to this resource is a very simple one and the search had to be modified accordingly:

(postnatal or antenatal or "post natal" or "ante natal" or postpartum or "postpartum" or pregnan* or PND or PPD) and depress*

*meta*Register of Controlled Trials (*m*RCT) (via Current Controlled Trials – http://controlled-trials.com/)

Searched 15 February 2007.

Retrieved 129 hits.

The search interface to this resource is a very simple one and the search had to be modified accordingly:

(Postnatal or "post natal" or "post natal" or prenatal or "pre-natal" or "pre natal" or perinatal or postpartum or "post partum" or "post-partum" or puerperal) AND depress% AND (diagnos% or screen% or detect% or predict% or identify%)

Health Services Research Projects in Progress (HSRProj) (www.nlm.nih.gov/hsrproj/)

Searched 15 February 2007.
Retrieved 24 hits.

The search interface to this resource is a very simple one and the search had to be modified accordingly:

(Postnatal OR "post natal" OR "post natal" OR prenatal OR "pre-natal" OR "pre natal" OR perinatal OR postpartum OR "post partum" OR "post-partum" OR puerperal) AND (depression OR depressive OR depressed)

LILACS (http://bases.bireme. br/cgibin/wxislind.exe/iah/ online/?IsisScript=iah/iah. xis&base=LILACS&lang=i)

Searched 15 February 2007.
Retrieved 94 hits.

The search interface to this resource is a very simple one and the search had to be modified accordingly:

Mh Pregnancy OR Mh Prenatal Care OR Mh Postnatal Care OR Tw pregnancy OR Tw pregnant OR Tw prenatal OR Tw pre-natal OR Tw postnatal OR tw postnatal OR Tw postpartum OR Tw post-partum OR Tw puerperal OR Tw new mother$OR Tw pre-pregnancy OR Tw prepregnancy OR Tw ante-natal OR Tw antenatal OR Tw antepartum OR Tw ante-partum

AND

Mh Depression OR Mh Depression, Postpartum OR Mh Depressive Disorder OR Mh anxiety OR Tw pnd OR Tw blues OR Tw depress$OR Tw melancholia OR Tw anxiety OR Tw anxious OR Tw ppd

AND

Tw screen$OR Tw diagnos$OR Tw detect$OR Tw predict$OR Tw aware$OR Tw identif$OR Mh diagnosis OR Mh questionnaires OR Mh interviews

Inside Conferences – Accessed via Dialog (file 65) using DialogLink 5

Searched 22 February 2007.

Retrieved 42 hits.

1. PREGNANCY/TI,AB,DE FROM 65
2. PREGNANT/TI,AB,DE FROM 65
3. PRENATAL/TI,AB,DE FROM 65
4. PRE(W)NATAL/TI,AB,DE FROM 65
5. POSTNATAL/TI,AB,DE FROM 65
6. POST(W)NATAL/TI,AB,DE FROM 65
7. POSTPARTUM/TI,AB,DE FROM 65
8. POST(W)PARTUM/TI,AB,DE FROM 65

165

9.   PUERPERAL/TI,AB,DE FROM 65
10. NEW(W)MOTHER?/TI,AB,DE FROM 65
11. PRE(W)PREGNANCY/TI,AB,DE FROM 65
12. PREPREGNANCY/TI,AB,DE FROM 65
13. ANTE(W)NATAL/TI,AB,DE FROM 65
14. ANTENATAL/TI,AB,DE FROM 65
15. ANTEPARTUM/TI,AB,DE FROM 65
16. ANTE(W)PARTUM/TI,AB,DE FROM 65
17. S1:S16 FROM 65
18. PND/TI,AB,DE FROM 65
19. BLUES/TI,AB,DE FROM 65
20. DEPRESS?/TI,AB,DE FROM 65
21. MELANCHOLIA/TI,AB,DE FROM 65
22. (ANXIETY OR ANXIOUS)/TI,AB,DE FROM 65
23. PPD/TI,AB,DE FROM 65
24. S18:S23 FROM 65
25. SCREEN?/TI,AB,DE FROM 65
26. DIAGNOS?/TI,AB,DE FROM 65
27. DETECT?/TI,AB,DE FROM 65
28. PREDICT?/TI,AB,DE FROM 65
29. AWARE?/TI,AB,DE FROM 65
30. IDENTIF?/TI,AB,DE FROM 65
31. DIAGNOSIS/TI,AB,DE FROM 65
32. (EDINBURGH 5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
33. EPDS/TI,AB,DE FROM 65
34. (POSTPARTUM (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
35. (POST-PARTUM (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/TI,AB,DE FROM 65
36. PDSS/TI,AB,DE FROM 65
37. (BROMLEY (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
38. BPDS/TI,AB,DE FROM 65
39. (GENERAL(W)HEALTH 5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
40. GHQ/TI,AB,DE FROM 65
41. (BECK (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
42. BDI/TI,AB,DE FROM 65
43. BAI/TI,AB,DE FROM 65
44. (STATE (2N) ANXIETY (2N) DEPRESSION)/ TI,AB,DE FROM 65

45. SAD/TI,AB,DE FROM 65
46. (HOSPITAL (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
47. HADS/TI,AB,DE FROM 65
48. (HAMILTON (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
49. HRSD/TI,AB,DE FROM 65
50. (ZUNG (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
51. SDS/TI,AB,DE FROM 65
52. PROFILE(2N)MOOD(W)STATES/TI,AB,DE FROM 65
53. POMS/TI,AB,DE FROM 65
54. ((CENTRE (2N) EPIDEMIOLOGICAL STUDIES) (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
55. (CES(W)D (2N) CESD)/TI,AB,DE FROM   65
56. SYMPTOM(W)CHECKLIST(W)90(W) REVISED/TI,AB,DE FROM 65
57. SCL(W)90(W)R/TI,AB,DE FROM 65
58. ((BRIEF(W)SYMPTOM) (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX   OR CHECKLIST OR INTERVIEW))/TI,AB,DE FROM 65
59. BSI/TI,AB,DE FROM 65
60. ((INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW) (5N) (DEPRESSIVE(W) SYMPTOMATOLOGY))/TI,AB,DE FROM 65
61. IDS/TI,AB,DE FROM 65
62. ((MONTGOMERY(W)ASBERG) (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/TI,AB,DE FROM 65
63. MADRS/TI,AB,DE FROM 65
64. ((DEPRESSIVE(W)ADJECTIVE) (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/TI,AB,DE FROM 65
65. DACL/TI,AB,DE FROM 65
66. (SCHEDULE (2N) AFFECTIVE(W) DISORDERS (2N) SCHIZOPHRENIA)/ TI,AB,DE FROM 65
67. SADS/TI,AB,DE FROM 65
68. ((STATE(W)TRAIT(W)ANXIETY) (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/TI,AB,DE FROM 65

69. STAI/TI,AB,DE FROM 65
70. ((BRISBANE(W)POSTNATAL) (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/TI,AB,DE FROM 65
71. ((POST(W)PARTUM(W)DEPRESSION(W) PREDICTORS) (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
72. ((POSTPARTUM(W)DEPRESSION(W) PREDICTORS) (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 65
73. ((DEPRESS? OR ANXIETY) (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/TI,AB,DE FROM 65
74. ANTENATAL(W)PSYCHOSOCIAL(W) HEALTH(W)ASSESSMENT/TI,AB,DE FROM 65
75. ALPHA/TI,AB,DE FROM 65
76. S25:S50 FROM 65
77. S51:S75 FROM 65
78. S76 OR S77 FROM 65
79. S17 AND S24 AND S78 FROM 65

Dissertation Abstracts – Accessed via Dialog (file 35) using DialogLink 5

Searched 22 February 2007.
Retrieved 740 hits.

1. PREGNANCY/TI,AB,DE FROM 35
2. PREGNANT/TI,AB,DE FROM 35
3. PRENATAL/TI,AB,DE FROM 35
4. PRE(W)NATAL/TI,AB,DE FROM 35
5. POSTNATAL/TI,AB,DE FROM 35
6. POST(W)NATAL/TI,AB,DE FROM 35
7. POSTPARTUM/TI,AB,DE FROM 35
8. POST(W)PARTUM/TI,AB,DE FROM 35
9. PUERPERAL/TI,AB,DE FROM 35
10. NEW(W)MOTHER?/TI,AB,DE FROM 35
11. PRE(W)PREGNANCY/TI,AB,DE FROM 35
12. PREPREGNANCY/TI,AB,DE FROM 35
13. ANTE(W)NATAL/TI,AB,DE FROM 35
14. ANTENATAL/TI,AB,DE FROM 35
15. ANTEPARTUM/TI,AB,DE FROM 35
16. ANTE(W)PARTUM/TI,AB,DE FROM 35
17. S1:S16 FROM 35
18. PND/TI,AB,DE FROM 35
19. BLUES/TI,AB,DE FROM 35
20. DEPRESS?/TI,AB,DE FROM 35
21. MELANCHOLIA/TI,AB,DE FROM 35
22. (ANXIETY OR ANXIOUS)/TI,AB,DE FROM 35
23. PPD/TI,AB,DE FROM 35
24. S18:S23 FROM 35
25. SCREEN?/TI,AB,DE FROM 35
26. DIAGNOS?/TI,AB,DE FROM 35
27. DETECT?/TI,AB,DE FROM 35
28. PREDICT?/TI,AB,DE FROM 35
29. AWARE?/TI,AB,DE FROM 35
30. IDENTIF?/TI,AB,DE FROM 35
31. DIAGNOSIS/TI,AB,DE FROM 35
32. (EDINBURGH (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 35
33. EPDS/TI,AB,DE FROM 35
34. (POSTPARTUM (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 35
35. (POST-PARTUM (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 35
36. PDSS/TI,AB,DE FROM 35
37. (BROMLEY (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 35
38. BPDS/TI,AB,DE FROM 35
39. (GENERAL(W)HEALTH (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 35
40. GHQ/TI,AB,DE FROM 35
41. (BECK (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 35
42. BDI/TI,AB,DE FROM 35
43. BAI/TI,AB,DE FROM 35
44. (STATE (2N) ANXIETY (2N) DEPRESSION)/ TI,AB,DE FROM 35
45. SAD/TI,AB,DE FROM 35
46. (HOSPITAL (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 35
47. HADS/TI,AB,DE FROM 35
48. (HAMILTON (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX OR CHECKLIST OR INTERVIEW))/ TI,AB,DE FROM 35
49. HRSD/TI,AB,DE FROM 35
50. (ZUNG (5N) (INVENTORY OR QUESTIONNAIRE OR SCALE OR INDEX

OR CHECKLIST OR INTERVIEW))/
TI,AB,DE FROM 35
51. SDS/TI,AB,DE FROM 35
52. PROFILE(2N)MOOD(W)STATES/TI,AB,DE
FROM 35
53. POMS/TI,AB,DE FROM 35
54. ((CENTRE (2N) EPIDEMIOLOGICAL
STUDIES) (5N) (INVENTORY OR
QUESTIONNAIRE OR SCALE OR INDEX
OR CHECKLIST OR INTERVIEW))/
TI,AB,DE FROM 35
55. (CES(W)D (2N) CESD)/TI,AB,DE FROM 35
56. SYMPTOM(W)CHECKLIST(W)90(W)
REVISED/TI,AB,DE FROM 35
57. SCL(W)90(W)R/TI,AB,DE FROM 35
58. ((BRIEF(W)SYMPTOM) (5N) (INVENTORY
OR QUESTIONNAIRE OR SCALE OR
INDEX OR CHECKLIST OR INTERVIEW))/
TI,AB,DE FROM 35
59. BSI/TI,AB,DE FROM 35
60. ((INVENTORY OR QUESTIONNAIRE
OR SCALE OR INDEX OR CHECKLIST
OR INTERVIEW) (5N) (DEPRESSIVE(W)
SYMPTOMATOLOGY))/TI,AB,DE FROM 35
61. IDS/TI,AB,DE FROM 35
62. ((MONTGOMERY(W)ASBERG) (5N)
(INVENTORY OR QUESTIONNAIRE OR
SCALE OR INDEX OR CHECKLIST OR
INTERVIEW))/TI,AB,DE FROM 35
63. MADRS/TI,AB,DE FROM 35
64. ((DEPRESSIVE(W)ADJECTIVE) (5N)
(INVENTORY OR QUESTIONNAIRE OR
SCALE OR INDEX OR CHECKLIST OR
INTERVIEW))/TI,AB,DE FROM 35
65. DACL/TI,AB,DE FROM 35
66. (SCHEDULE (2N) AFFECTIVE(W)
DISORDERS (2N) SCHIZOPHRENIA)/
TI,AB,DE FROM 35
67. SADS/TI,AB,DE FROM 35
68. ((STATE(W)TRAIT(W)ANXIETY) (5N)
(INVENTORY OR QUESTIONNAIRE OR
SCALE OR INDEX OR CHECKLIST OR
INTERVIEW))/TI,AB,DE FROM 35
69. STAI/TI,AB,DE FROM 35
70. ((BRISBANE(W)POSTNATAL) (5N)
(INVENTORY OR QUESTIONNAIRE OR
SCALE OR INDEX OR CHECKLIST OR
INTERVIEW))/TI,AB,DE FROM 35
71. ((POST(W)PARTUM(W)DEPRESSION(W)
PREDICTORS) (5N) (INVENTORY OR
QUESTIONNAIRE OR SCALE OR INDEX
OR CHECKLIST OR INTERVIEW))/
TI,AB,DE FROM 35
72. ((POSTPARTUM(W)DEPRESSION(W)
PREDICTORS) (5N) (INVENTORY OR
QUESTIONNAIRE OR SCALE OR INDEX

OR CHECKLIST OR INTERVIEW))/
TI,AB,DE FROM 35
73. ((DEPRESS? OR ANXIETY) (5N)
(INVENTORY OR QUESTIONNAIRE OR
SCALE OR INDEX OR CHECKLIST OR
INTERVIEW))/TI,AB,DE FROM 35
74. ANTENATAL(W)PSYCHOSOCIAL(W)
HEALTH(W)ASSESSMENT/TI,AB,DE FROM
35
75. ALPHA/TI,AB,DE FROM 35
76. S25:S50 FROM 35
77. S51:S75 FROM 35
78. S76 OR S77 FROM 35
79. S17 AND S24 AND S78 FROM 35

Economic databases
*NHS Economic Evaluation Database
(NHS EED) (CRD – www.york.ac.uk/inst/
crd/crddatabases.htm) April 2007 update*

Searched 21 May 2007.
Retrieved 402 hits.

#1. MeSH Depression
#2. pnd
#3. blues
#4. Depress*
#5. melancholia
#6. MeSH Depressive Disorder EXPLODE 1
#7. ppd
#8. "Seasonal mood disorder*"
#9. Sad
#10. "Seasonal affective disorder*"
#11. anxiety OR anxious
#12. MeSH Anxiety
#13. #1 or #2 or #3 or #4 or #5 or #6 or #7 or
#8 or #9 or #10 or #11 or #12
#14. "economic evaluation":ty OR "provisional
abstract":ty
#15. #13 and #14

*Health Economic Evaluations Database
(HEED) (CD-ROM) April 2007 update*

Searched 21 May 2007.
Retrieved 706 hits.

1. ti=pnd or ab=pnd or kw=pnd
2. ti=blues or ab=blues or kw=blues
3. ti=depress* or ab=depress* or kw=depress*
4. ti=melancholia or ab=melancholia or
kw=melancholia
5. ti=ppd or ab=ppd or kw=ppd
6. ti='seasonal mood disorder*' or ab='seasonal
mood disorder*' or kw='seasonal mood
disorder*'

7.   ti=sad or ab=sad or kw=sad
8.   ti='seasonal affective disorder' or ab='seasonal affective disorder*' or kw='seasonal affective disorder*'
9.   ti=anxiety or ab=anxiety or kw=anxiety
10. ti=anxious or ab=anxious or kw=anxious
11. ic=300.4 or ic=296.2 or ic=298.0 or ic=296.3 or ic=311
12. cs= 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11

IDEAS (http://ideas.repec.org/)

Searched 21 May 2007.
Retrieved 165 hits.
Used detailed search screen:
Match: Boolean
Keyword precision: fuzzy

pnd or blues or depress* or melancholia or ppd or (seasonal and mood and disorder*) or (seasonal and affective and disorder*) or sad or anxiety or anxious

EconLit (ERLWebSPIRS5 – http://arc. uk.ovid.com/) 1969 to April 2007

Searched on 22 May 2007.
Retrieved 2780 hits.

#1.   DEPRESSION-MENTAL-HEALTH in DES
#2.   ((pnd) in AB)or((pnd) in TI)
#3.   ((blues) in AB)or((blues) in TI)
#4.   ((Depress*) in AB)or((Depress*) in TI)
#5.   ((melancholia) in AB)or((melancholia) in TI)
#6.   ((ppd) in AB)or((ppd) in TI)
#7.   ((Seasonal mood disorder*) in AB) or((Seasonal mood disorder*) in TI)
#8.   ((Sad) in AB)or((Sad) in TI)
#9.   ((Seasonal affective disorder*) in AB) or((Seasonal affective disorder*) in TI)
#10. ((anxiety) in TI)or((anxiety) in AB)
#11. ((anxious) in AB)or((anxious) in TI)
#12. #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11

Appendix 2
Stakeholder involvement

Stakeholder invitees

Professor Peter Brocklehurst	Professor of Perinatal Epidemiology
Mrs Paula Dunn	Service user
Mrs Tracy Lamb	Service user
Mrs Cheryll Adams	Acting PO of Unite/Community Practitioners' and Health Visitors' Association
Mrs Kay Bennett	Community Midwife
Dr Roch Cantwell	Psychiatrist
Mrs Jan Cubison	Maternal Mental Health Service Co-ordinator
Mrs Sandra Elliott	Consultant Clinical Psychologist
Mrs Debbie Fielding	Primary care trust commissioner for women's and children's services
Dr Ron Gray	Psychiatrist
Professor Josephine Green	Professor of Psychosocial Reproductive Health
Mrs Julia Hanrahan	Community Psychiatric Nurse
Mrs Roslyn Hope	Director of National Workforce Programme
Mrs Mervi Jokinen	Practice and Standards Development Advisor for the Royal College of Midwives
Professor Helen Lester	Professor of Primary Care
Mrs Catherine Lowenhoff	Nurse Advisor to the Department for Children's Schools and Families
Mrs Karen Newbigging	Joint National Lead for Gender Equality and Women's Mental Health
Dr Margaret Oates	Psychiatrist
Mrs Margeret Petty	Health Visitor
Mrs Karen Robertson	Nurse Consultant in Perinatal Mental Health
Mrs Ruth Rothman	Specialist Lead for Parental and Child Mental Health
Mrs Sally Russell	Service user
Dr Sheelah Seeley	Health Visitor Training Facilitator
Dr James Seward	National Programme Director for Care Services Improvement Partnership
Dr Judy Shakespeare	General Practitioner
Dr Dave Tomson	General Practitioner
Mrs Rosie Jones	Health Visitor
Dr Dick Churchill	Senior Lecturer
Mr Barry Nixon	National Lead for Child and Adolescent Mental Health in the National Workforce Programme
Mrs Jan Keithson	Social Worker

Stakeholder attendees

Mrs Ruth Rothman	Specialist Lead for Parental and Child Mental Health
Mrs Catherine Lowenhoff	Nurse Advisor to the Department for Children's Schools and Families
Dr Judy Shakespeare	General Practitioner
Professor Josephine Green	Professor of Psychosocial Reproductive Health
Professor Peter Brocklehurst	Professor of Perinatal Epidemiology
Mrs Margeret Petty	Health Visitor
Mrs Paula Dunn	Service user
Mr Barry Nixon	National Lead for Child and Adolescent Mental Health in the National Workforce Programme
Mrs Rosie Jones	Health Visitor
Mrs Jan Keithson	Social Worker
Mrs Tracy Lamb	Service user

Appendix 3

Youden's index

Reference test		
Index test	+	–
+	True positive[a]	False positive[c]
–	False negative[b]	True negative[d]

a people with disease identified by the index test (true positive)
b people with disease not identified by the index test (false negative)
c people without disease identified by the index test (false positive)
d people without disease not identified by the index test (true negative)
a + b = people with disease
c + d = people without disease

Youden's index

$$= \frac{a}{a+b} + \frac{d}{c+d} - 1 = \text{(sensitivity)} + \text{(specificity)} - 1$$

Youden's index ranges from –1 to +1, with a perfect test having a value of +1.

Example

Reference test		
Index test	+	–
+	50	0
–	15	35

Youden's index

$$= \frac{50}{50+15} + \frac{35}{0+35} - 1 = \frac{50}{65} + \frac{35}{35} - 1 = 0.77 + 1 - 1 = 0.77$$

Appendix 4

sROC plots

sROC plots for major depression only
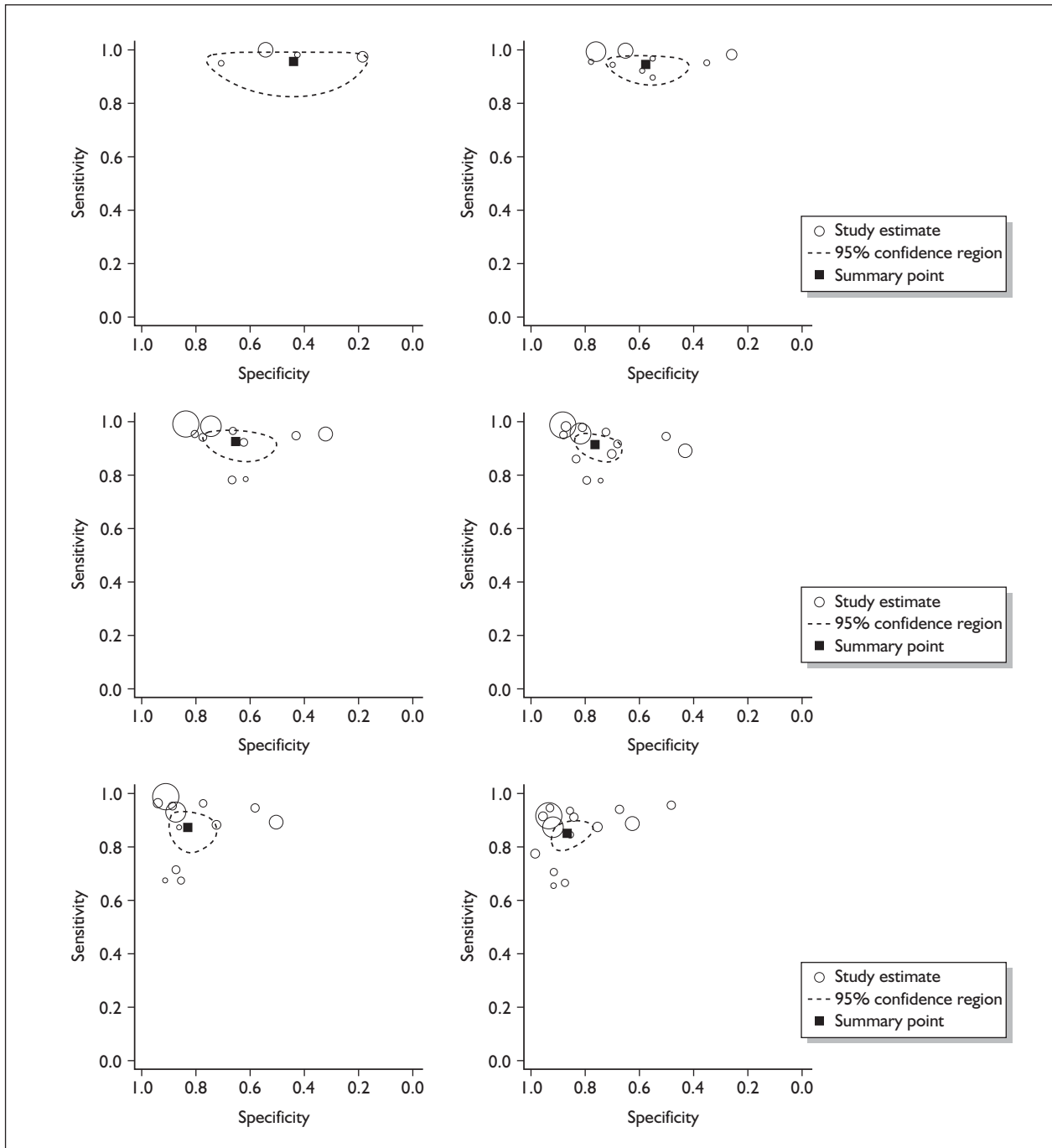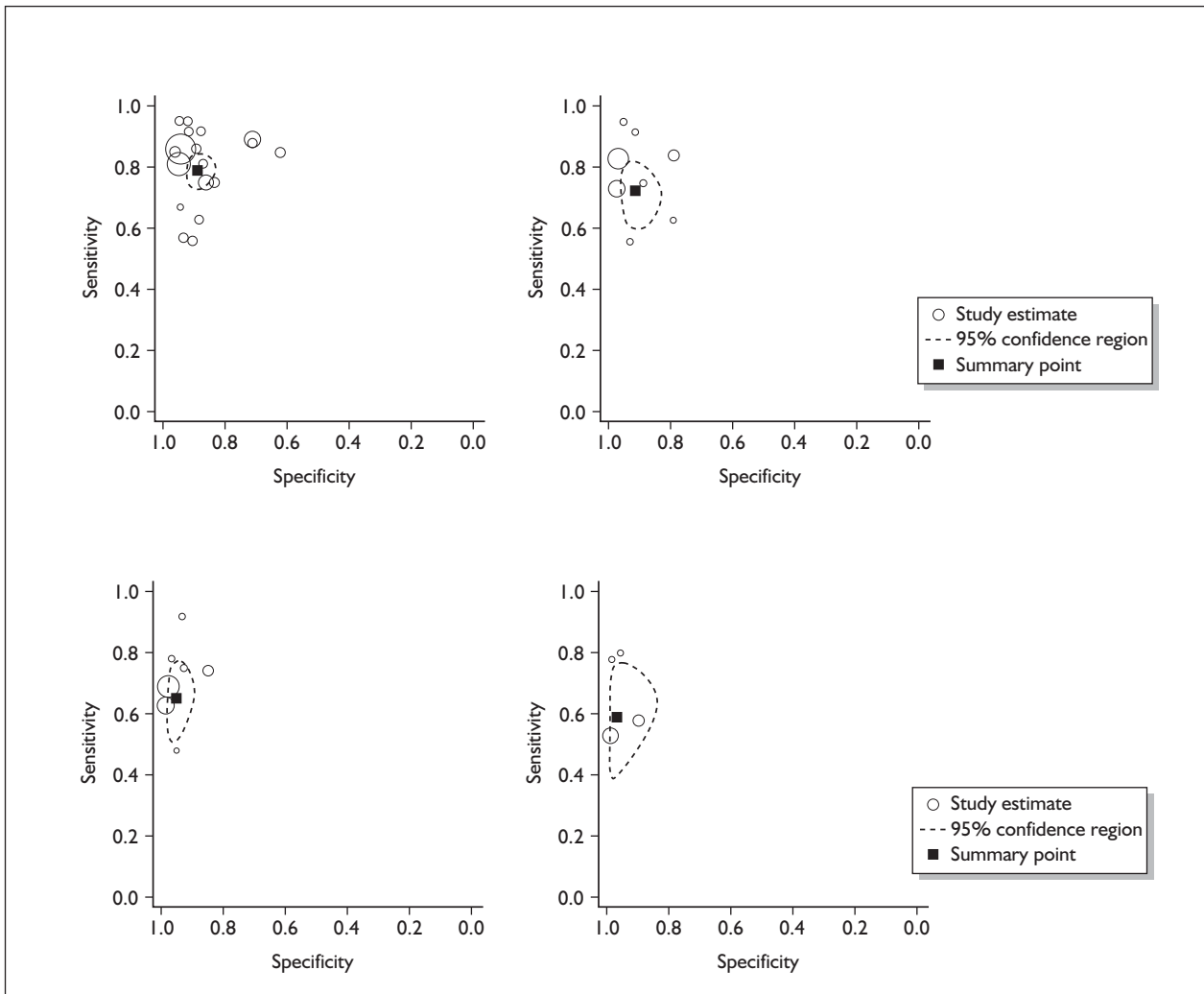
sROC plots for major or minor depression
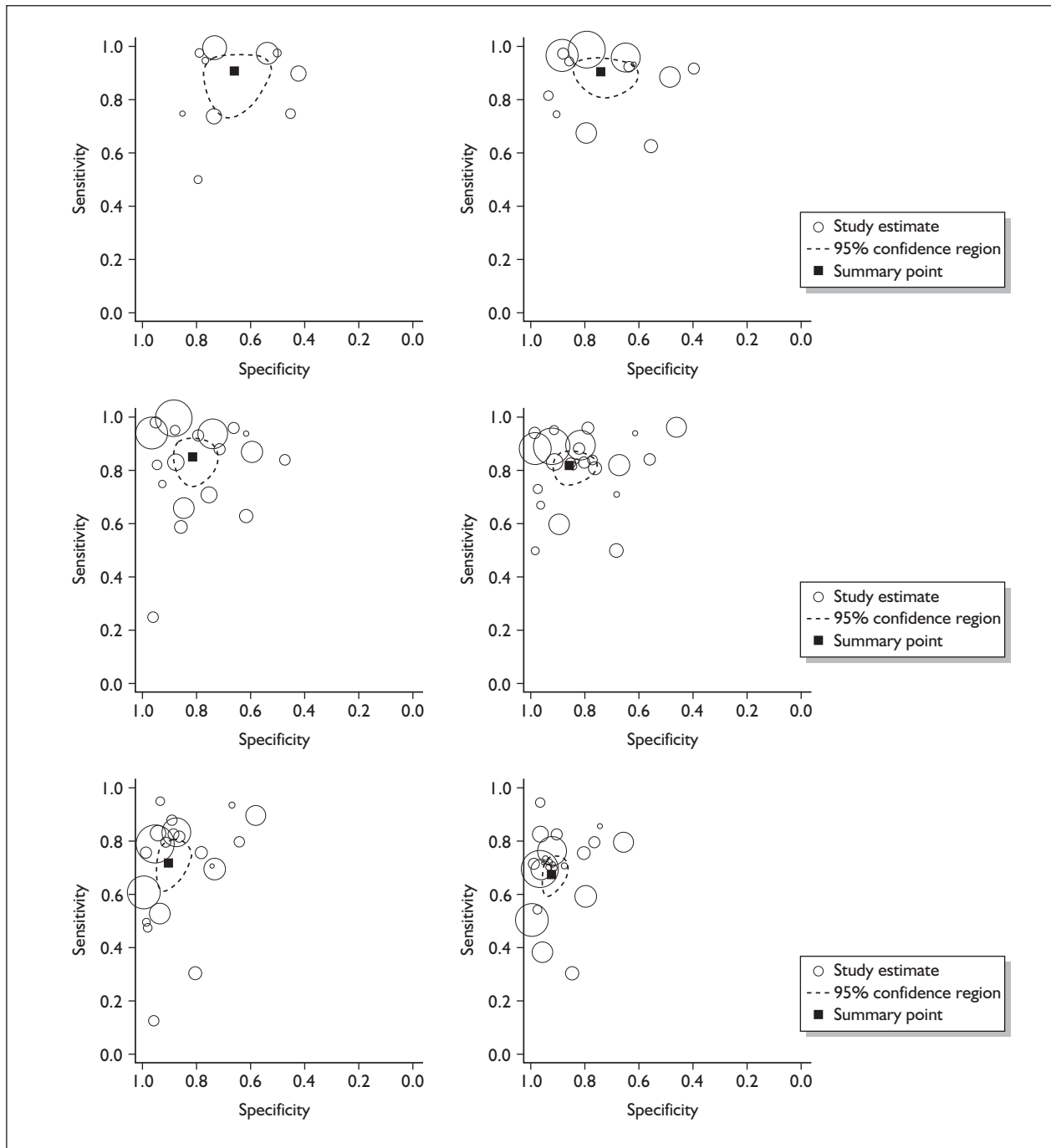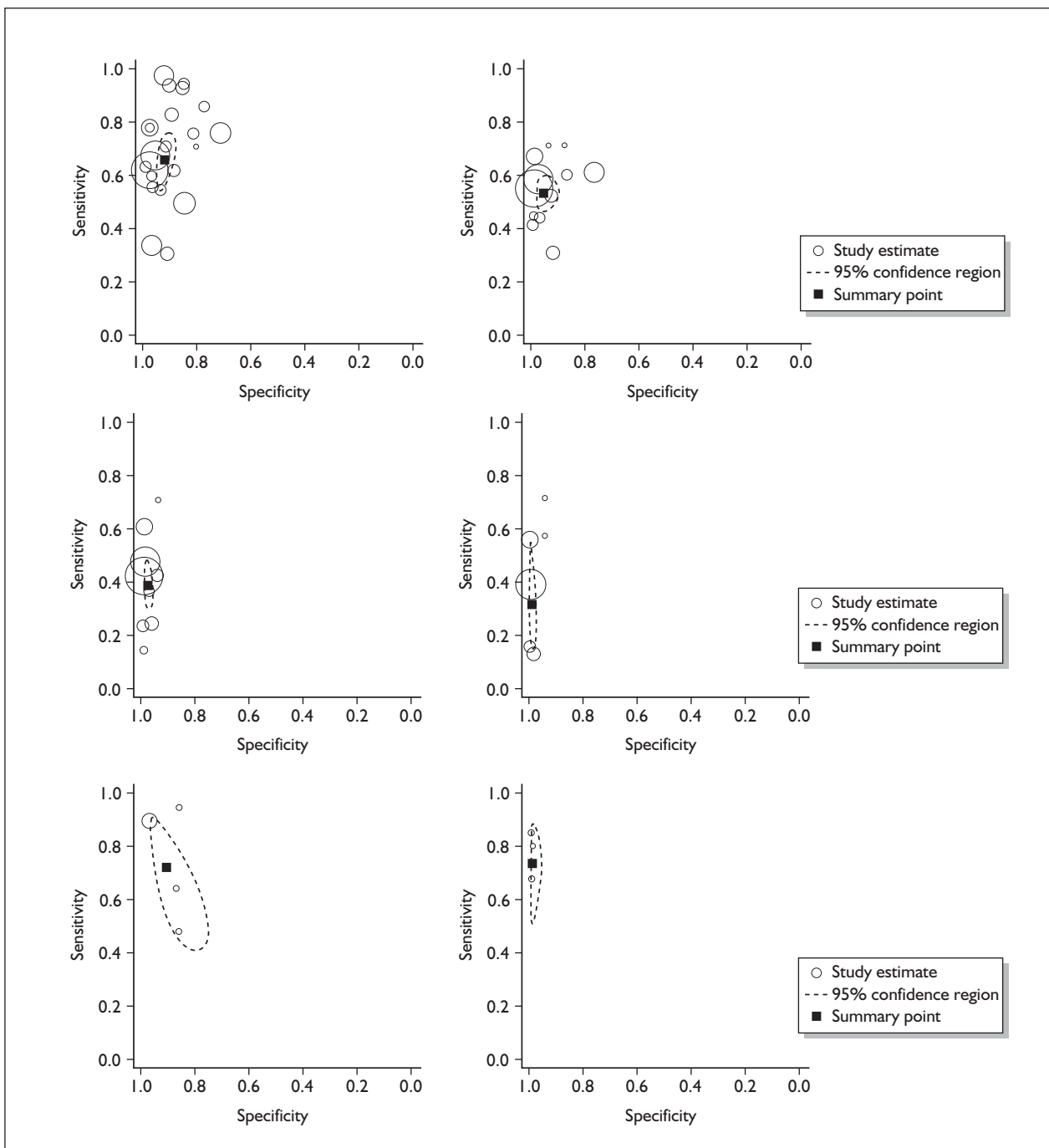
sROC plots for any psychiatric disorder
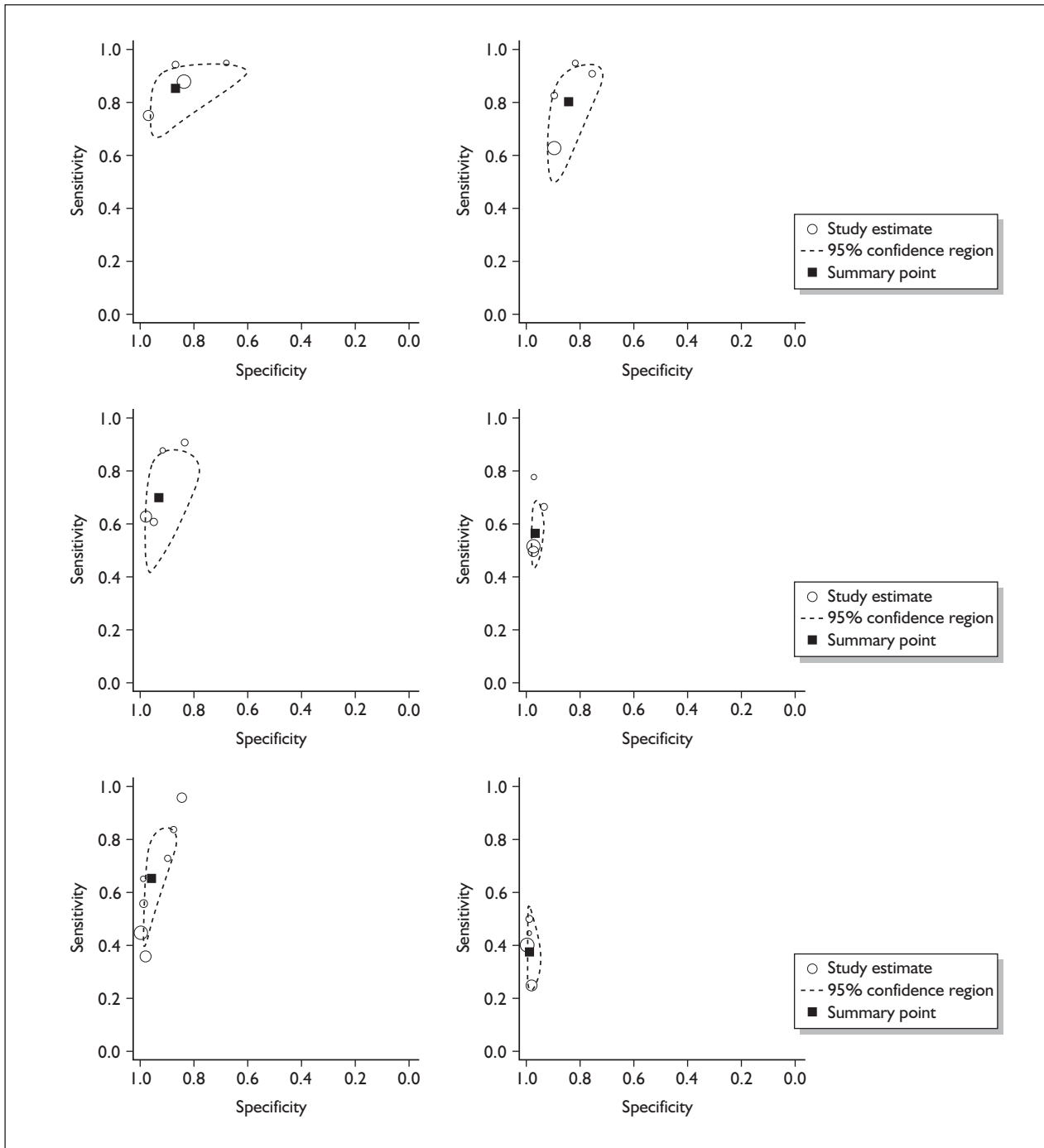
Appendix 5

Summary table for review 3

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Evins et al., 2000,[191] USA Individual RCT (odd/even dates of delivery) 391 women	Hospital for recruitment. EPDS by mail Consenting women who delivered a baby at the hospital	**Intervention:** Screening for PND with the EPDS (postal). Women at high risk (EPDS ≥ 10) for PND were referred to the social worker for assessment **Usual care:** No screening with the EPDS, only spontaneous detection during routine clinical evaluation	**Threshold EPDS scores at 6 weeks:** 28 (35.4%) of the 79 women in the intervention group had EPDS score ≥ 10 vs 6 (6.3%) of the 96 women in the control group were identified as having PND during routine clinical evaluation ($p < 0.0001$)
Kung et al., 2002,[194] China Individual RCT 405 women	Hospital for recruitment. Home for the intervention Chinese women who were Hong Kong residents, who spoke Cantonese and who gave birth to a baby in a single hospital were included	**Intervention:** Early intervention including in-depth assessment by midwives (women were asked to describe any sleep disturbance, poor appetite, self-blame or if they had ideas of hurting the baby) and subsequent phone follow-up service by midwives and volunteers to the women with high EPDS scores. Therapeutic listening, information and appropriate care were provided according to need. If women were in severe distress they were referred to an obstetrician. Information on community resources were also offered if required. A contact telephone number was given to anyone who refused weekly telephone follow-up so that they could contact the midwife if they needed help. **Enhancement:** Training on basic counselling and interviewing skills and emotional problems of postnatal women were provided to midwives, social workers and volunteers before the study **Usual care:** Women were not assessed with the EPDS until 6 weeks after delivery. Emotional state was assessed by routine clinical evaluation and counselling provided if needed. A contact telephone number was also given to women for them to call back if they needed help or wanted to talk with the midwife	**Threshold EPDS scores at 6 weeks:** 12 (5.9%) of the 203 women in the intervention group vs 24 (11.8%) of the 202 women in the control group had EPDS score > 9 ($p = 0.03$). Mean EPDS scores at 6 weeks: The mean score of the intervention group was 4.32 whereas that of the control group was 4.97 ($p = ns$). Use of contact number: Four women in the control group used the contact telephone number
MacArthur et al., 2002,[193] UK Cluster RCT 36 GP practices	Women who received postnatal care in the recruited practices	**Intervention:** Care was led by midwives, with contact with GPs based on referral, including home visits and the final discharge consultation. A symptoms checklist was used at days 10 and 28 and at the discharge consultation at 10–12 weeks. The EPDS was also used at day 28 and discharge consultation. Enhancement: Midwives trained in postnatal care and health and trial design. Training to implement new model of care **Usual care:** Standard community postnatal care consisting of seven home midwife visits up to 10–14 days (can continue to day 28) after birth and care from health visitors thereafter. GPs did routine home visits and final 6- to 8-week check. Some health visitors use the EPDS routinely so some control subject screened. **Enhancement:** Midwives trained in postnatal care and health and trial design	**Mean mental health score at 16 weeks:** mean of cluster mean = 50.5 vs 47.54, difference = 2.96 (95% CI 1.16 to 4.77), $p = 0.002$; multilevel OR = 3.03 (95% CI 1.53 to 4.52), $p < 0.0001$; OR_{adj} = 4.31 (95% CI 2.50 to 6.12). **Mean EPDS score at 16 weeks:** mean of cluster mean = 6.40 vs 8.06, difference = −1.66 (95% CI −2.49 to −0.83), $p < 0.0001$; multilevel OR = −1.92 (95% CI −2.55 to −1.29), $p < 0.0001$; OR_{adj} = −2.68 (95% CI −3.46 to −1.89). **Mean physical health score at 16 weeks:** mean of cluster mean = 46.68 vs 47.84, difference = −1.17 (95% CI −2.52 to 0.19), $p = 0.09$; multilevel OR = −0.79 (95% CI −1.91 to 0.34), $p = 0.17$; OR_{adj} = −0.80 (95% CI −2.32 to 0.72).

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			Threshold EPDS scores at 16 weeks postnatally: percentage of cluster = 14.4 vs 21.3, difference = −6.85% (95% CI −11.99 to −1.71), $p = 0.01$; multilevel OR = 0.57 (95% CI 0.43 to 0.76), $p = 0.0001$; OR$_{adj}$ = 0.47 (95% CI 0.31 to 0.76). **Overall satisfaction:** OR = 1.09, 95% CI 0.72 to 1.63, $p = 0.7$. **Better than expected:** OR = 1.35, 95% CI 1.08 to 1.70, $p = 0.01$. **Planning care score** coefficient = 0.49, 95% CI 0.13 to 0.85, $p = 0.01$. **Continuity of care** coefficient = 0.21, 95% CI −0.11 to 0.52, $p = 0.2$. **Maternity discharge score** coefficient = 0.19, 95% CI −0.11 to 0.50, $p = 0.2$. **Talk to midwife about most/all health symptoms:** OR = 1.52, 95% CI 1.05 to 2.20, $p = 0.03$. **No difficulty talking to midwife about health symptoms:** OR = 1.61, 95% CI 1.07 to 2.41, $p = 0.02$. **Midwife visits (from midwife records):** mean of cluster mean: 6.00 (1.17) vs 4.07 (1.37), difference 1.92 (95% CI 1.04 to 2.80), $p < 0.0001$. **Midwife visits (from women's diaries):** mean of cluster mean: 5.82 (1.09) vs 5.99 (1.02), difference −0.17 (95% CI −0.89 to 0.54), $p = 0.63$. **GP visits (from 4-month questionnaire):** mean of cluster mean: 0.60 (0.38) vs 0.78 (0.55), difference −0.18 (95% CI −0.51 to 0.14), $p = 0.26$. **GP visits (from women's diaries):** mean of cluster mean: 0.57 (0.37) vs 0.88 (0.63), difference −0.31 (95% CI −0.67 to 0.04), $p = 0.08$
Webster *et al.*, 2003,[195] Australia Individual RCT 600 women	Obstetric (prenatal clinic) hospital Risk factors for PND: (1) low social or partner support, measured by Maternity Social Support Scale ≤ 24, (2) a past history of mental illness, (3) family psychiatric history, (4) past PND or (5) having a mother who had PND. Only positively screened women were randomly allocated	**Intervention:** Educational intervention, providing women with a booklet about PND and a list of phone contacts of PND resources prenatally. Letters were also sent to referring GP and to their child health nurse with details of their risk status **Usual care:** Standard care, including case management and referral to a hospital social worker or psychiatrist if required	**Threshold EPDS scores at 16 weeks:** 46 (24%) of the 192 women in the intervention group vs 50 (24%) of the 177 women in the control group had EPDS score > 12, OR = 0.80, 95% CI 0.50 to 1.28

continued

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Wickberg et al., 2005,[192] Sweden Cluster controlled clinical trial (midwives allocated based on sectors) 32 midwives, 669 women	Antenatal care centres in six sectors in Sweden Ability to understand Swedish	**Intervention:** Feedback of EPDS score if 12 or above at 25 weeks' gestation. Enhancement of care: Midwives in the study group received a one-afternoon information session about different aspects of depression, such as symptoms, aetiology and effects, and about the value of listening and support **Usual care:** Not described	**Threshold EPDS scores at gestational week 25:** 48 (15%) of the 318 women in the intervention group vs 45 (13%) of the 351 women in the control group had EPDS score > 11. **Threshold EPDS scores at gestational week 36:** 26 (10%) of the 273 women in the intervention group vs 40 (12%) of the 345 women in the control group had EPDS score > 11; this difference was statistically significant ($p < 0.0001$). **Mean EPDS scores at gestational week 25:** 6.41 (range 0–25) in the study group and 6.07 (range 0–21) in the control group, $p = 0.54$. **Mean EPDS scores at gestational week 36:** 5.39 (range 0–19) in the study group and 6.11 (range 0–22) in the control group, $p = 0.12$. **Within-group changes:** The within-group mean difference in EPDS scores was statistically significant ($p = 0.0003$) for the study group but not for the control group. **EPDS change score:** EPDS score decreased by 1 point in the study group (mean value = –1.02) between weeks 25 and 36 and was almost the same in the comparison group (mean value = 0.04). This difference between groups was significant (U = 22,544, $p < 0.05$). **Health care:** Among 77 high-scoring women in the study who completed a questionnaire at the postnatal visit the health care received was similar in the two groups with no significant differences in number of visits to the doctor, number of visits to the midwife, sufficient number of visits to the midwife and number of midwives met during pregnancy. Nor did the number of referrals to a mental health professional differ significantly, but more women were referred in the study group (13/34) than in the comparison group (5/31)
Elliot et al., 1988,[98] UK Individual controlled trial 99 women	All women attending Lewisham Hospital obstetric department Women who attended first hospital appointment before the 18th week of pregnancy, spoke English and would not be moving from the area and who were not psychotic, not addicted to drugs, over 18 years, married, cohabiting or in a stable relationship and vulnerable to PND (Leverton questionnaire – either poor marital relationship or previous psychological problems or CCEI ≥ 10)	**Intervention:** Monthly meetings at hospital prenatal clinic from approximately 4 months of pregnancy to 6 months postnatally, 11 sessions in all. The programme included continuity of care, an educational component and sources of information. Health visitors also made additional visits early in pregnancy to enhance continuity of care and provide easier access to professional support **Usual care:** Unclear	**Bedford College criteria at 2 months:** 6 (12.5%) of the 48 women in the intervention group vs 17 (33.3%) of the 51 women in the control group received a diagnosis of depression, $p < 0.02$. **Bedford College criteria at 3 months:** 4 (8.3%) of the 48 women in the intervention group vs 8 (15.7%) of the 51 women in the control group received a diagnosis of depression, p = not significant. Further outcome data were reported for first-time and second-time mothers separately but are not reported here

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Holden *et al.*, 1989,[218] UK Individual randomised controlled trial 55 women	Women attending child health clinics at five centres EPDS score ≥ 13 at 6 weeks postnatally	**Intervention:** Eight weekly counselling visits by health visitors using Rogerian or non-directive counselling methods. Enhancements: Health visitors given a short training session in counselling for PND **Usual care:** Not described	**RDC at 3 months:** 8 (31%) of the 26 women in the intervention group met RDC criteria for major or minor depression vs 15 (62%) of the 24 in the control group, $p = 0.03$. **Within-group change in standardised psychiatric interview total score:** at time 1 median score was 25.5 and at time 2 median score was 14.0 in the intervention group ($n = 26$), median change = −12.5, 95% CI −17 to −7, $p = 0.001$; at time 1 median score was 24.0 and at time 2 median score was 23.0 in the control group ($n = 24$), median change = −2.0, 95% CI −6 to 3, $p = $ ns; significant difference in changes between groups ($p = 0.01$). **Within-group change in standardised psychiatric interview observed depression:** at time 1 median score was 2 and at time 2 median score was 0.5 in the intervention group ($n = 26$), median change = −2.0, 95% CI −2 to −1, $p = 0.001$; at time 1 median score was 2.0 and at time 2 median score was 2 in the control group ($n = 24$), median change = 0, 95% CI −1 to 0, $p = $ ns; significant difference in changes between groups ($p = 0.01$). **Within-group change in EPDS:** at time 1 median score was 16.0 and at time 2 median score was 10.5 in the intervention group ($n = 26$), median change = −6.0, 95% CI −8 to −4.5, $p = 0.001$; at time 1 median score was 15.5 and at time 2 median score was 12.0 in the control group ($n = 24$), median change = −1.5, 95% CI −4 to 0.5, $p = $ ns; significant difference in changes between groups ($p = 0.01$)
Stamp *et al.*, 1995,[203] Australia Individual RCT 144 women vulnerable for PND screened	Women's and children's hospital for high-risk pregnancies English-speaking women with a singleton pregnancy of less than 24 weeks' gestation, who lived within the metropolitan area and agreed to attend extra groups if invited, were eligible to enter the trial if they scored 2 or more on the Modified Antenatal Screening Questionnaire (MASQ)	**Intervention:** Routine care plus two special antenatal groups (32 and 36 weeks' gestation) and a postnatal group (6 weeks postnatally). The classes focused on access to information, preparation and support, the extension and development of women's existing networks and goal setting **Usual care:** Routine care including antenatal classes offered by the hospital, which at the time did not include specific information about PND until 6 weeks postnatally when a video was shown	**Threshold EPDS at 6 weeks:** 8 (13%) of 64 women in the intervention group had EPDS score > 12 vs 11 (17%) of 64 control women, OR = 0.69, 95% CI 0.22 to 2.03; 22 (34%) of 64 women in the intervention group had EPDS score > 9 vs 22 (34%) of 64 control women, OR = 1.00, 95% CI 0.45 to 2.21. **Threshold EPDS at 12 weeks:** 7 (11%) of 63 women in the intervention group had EPDS score > 12 vs 10 (15%) of 65 control women, OR = 0.69, 95% CI 0.22 to 2.14; 14 (22%) of 63 women in the intervention group had EPDS score > 9 vs 17 (26%) of 65 control women, OR = 0.81, 95% CI 0.33 to 1.96. **Threshold EPDS at 6 months:** 9 (15%) of 60 women in the intervention group had EPDS score > 12 vs 6 (10%) of 61 control women, OR = 1.62, 95% CI 0.47 to 5.91; 14 (23%) of 60 women in the intervention group had EPDS score > 9 vs 10 (16%) of 61 control women, OR = 1.55, 95% CI 0.58 to 4.22

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Meager and Milgrom, 1996,[220] Australia Individual RCT (pilot study) 20 women	Women recruited through adverts in local hospitals and maternal and child health centres Women included had developed their depressive condition within 6 months postnatally, had a rating scale of > 12 on the EPDS and a BDI score > 15	**Intervention:** 10-week group treatment programme with educational, social support and cognitive behavioural components **Usual care:** Waiting list control group	**Mean BDI:** week 0 = 29.7 (n = 10) and week 10 = 16.8 (n = 6) in the intervention group vs week 0 = 29.0 (n = 10) and week 10 = 29.1 (n = 6) in the control group; within intervention group p < 0.05 and difference between groups at week 10 p < 0.10. **Mean EPDS:** week 0 = 24.8 (n = 10) and week 10 = 15.8 (n = 6) in the intervention group vs week 0 = 27.5 (n = 10) and week 10 = 28.0 (n = 6) in the control group; within intervention group p < 0.02 and difference between groups at week 10 p < 0.05. **Mean POMS depression:** week 0 = 30.8 (n = 10) and week 10 = 13.6 (n = 6) in the intervention group vs week 0 = 33.8 (n = 10) and week 10 = 36.1 (n = 6) in the control group; within intervention group p < 0.01 and difference between groups at week 10 p < 0.02. **Mean POMS tension:** week 0 = 19.0 (n = 10) and week 10 = 9.80 (n = 6) in the intervention group vs week 0 = 22.4 (n = 10) and week 10 = 23.0 (n = 6) in the control group; within intervention group p < 0.02 and difference between groups at week 10 p < 0.02. **Mean POMS confusion:** week 0 = 16.5 (n = 10) and week 10 = 8.0 (n = 6) in the intervention group vs week 0 = 16.7 (n = 10) and week 10 = 20.1 (n = 6) in the control group; within intervention group p < 0.01 and difference between groups at week 10 p < 0.01. **Mean POMS fatigue:** week 0 = 21.0 (n = 10) and week 10 = 11.2 (n = 6) in the intervention group vs week 0 = 19.2 (n = 10) and week 10 = 19.3 (n = 6) in the control group; within intervention group p < 0.01 and difference between groups at week 10 p < 0.05. **Mean Parenting Stress Index (PSI):** week 0 = 313.4 and week 10 = 295.4 in the intervention group vs week 0 = 318.1 and week 10 = 341.4 in the control group. **Child domain subscale of PSI:** the intervention and control group mean scores at the start of the programme were 131.2 (n = 10) and 130.1 (n = 10), respectively, whereas at the end of the programme the intervention group mean score was 127.4 (n = 6) and the control group mean score was 146.0 (n = 6). **Dyadic Adjustment Scale, Coopersmith self-esteem measure and Social Provision Scale:** no significant differences were found on these measures

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Wickberg and Hwang, 1996,[219] Sweden Individual RCT 48 women	17 child health clinics Swedish-speaking mothers completed EPDS at 2 and 3 months postnatally. Women scoring ≥ 12 on both occasions were assessed using MADRS and DSM-III-R at home. All women with a MADRS score of ≥ 10 were invited to take part in the study	**Intervention:** Routine care plus six weekly 1-hour counselling sessions by the nurse in the home or at the clinics. Enhancement: Four half-day training sessions in non-directive counselling methods **Usual care:** Routine care, not consisting of any scheduled check-ups but the possibility of visiting the clinic whenever needed	**DSM-III-R criteria post intervention:** 3 (15%) of the 20 women in the intervention group vs 12 (57%) of the 21 women in the control group met DSM-III-R criteria for major depression, $p < 0.01$. **Within-group change in mean MADRS score:** baseline = –19.6 vs post intervention = –10.9 ($p = 0.0005$) in the intervention group; baseline = –17.1 vs post intervention = –14.7 (p = not significant) in the control group. **Well-being (self-complete):** in the intervention group 17 women felt much better or better; 3 felt neither better nor worse and none felt worse vs in the control group 12 women felt much better or better, 6 felt neither better nor worse and 3 felt much worse or worse. Further outcome data were reported regarding perceptions of the intervention but are not reported here
Buist *et al.*, 1999,[197] Australia Individual RCT 44 women	Antenatal clinics at a major obstetric teaching hospital Primiparous English-speaking women between 12 and 24 weeks' gestation with a score of ≥ 8 on the risk factors scale (score reflected a mix of three or more of a family history or past history of depression, PMT, marital or childhood difficulties)	**Intervention:** 10 intervention classes with partners, two of which were held postnatally, run by a midwife and a psychologist with experience of antenatal education and/or working in a mother baby unit. The sessions covered the physical preparation for parenthood but focused on emotional issues and highlighted the reality of parenting. Didactic teaching was combined with interactive group work, films and experimental exercises **Usual care:** Six standard antenatal classes	**Mean EPDS at 6 weeks:** 7.4 intervention ($n = 20$) vs 9.1 control ($n = 16$). **Mean EPDS at 6 months:** 7.6 intervention ($n = 20$) vs 8.1 control ($n = 16$). **Mean BDI at 6 weeks:** 7.8 intervention ($n = 19$) vs 11.2 control ($n = 16$). **Mean BDI at 6 months:** 7.8 intervention ($n = 12$) vs 9.4 control ($n = 10$). **Mean SAS STATS at 6 weeks:** 31.9 intervention ($n = 16$) vs 38.2 control ($n = 10$). **Mean SAS STATS at 6 months:** 32.7 intervention ($n = 12$) vs 36.4 control ($n = 10$); **Mean SAS TRAIT at 6 weeks:** 33.6 intervention ($n = 14$) vs 42.7 control ($n = 7$). **Mean SAS TRAIT at 6 months:** 37.6 intervention ($n = 9$) vs 43.0 control ($n = 3$). **Mean SDA at 6 weeks:** 30.7 intervention ($n = 19$) vs 29.1 control ($n = 16$). **Mean SDA at 6 months:** 27.3 intervention ($n = 13$) vs 29.4 control ($n = 10$). **Mean SSS number at 6 weeks:** 4.5 intervention ($n = 19$) vs 2.9 control ($n = 14$). **Mean SSS number at 6 months:** 3.7 intervention ($n = 14$) vs 1.6 control ($n = 10$), $p < 0.001$ (different at baseline). **Mean SSS satisfaction at 6 weeks:** 5.6 intervention ($n = 19$) vs 4.9 control ($n = 14$), $p < 0.05$. **Mean SSS satisfaction at 6 months:** 5.0 intervention ($n = 14$) vs 3.3 control ($n = 10$), $p < 0.001$ (different at baseline). **Mean group change preclass to 6-week BDI:** 9.6 to 7.9 in the intervention group ($n = 18$) vs 9.6 to 11.1 in the control group ($n = 16$).

187

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			Mean group change preclass to 6-month BDI: 9.1 to 8.2 in the intervention group ($n = 11$) vs 8.5 to 9.4 in the control group ($n = 10$). **Mean group change preclass to 6-week SAS STATS:** 35.5 to 32.0 in the intervention group ($n = 15$) vs 38.0 to 38.2 in the control group ($n = 10$). **Mean group change preclass to 6-month SAS STATS:** 36.1 to 33.4 in the intervention group ($n = 11$) vs 39.4 to 36.4 in the control group ($n = 10$). **Mean group change preclass to 6-week SAS TRAIT:** 41.1 to 33.6 in the intervention group ($n = 14$) ($p < 0.05$) vs 44.2 to 42.7 in the control group ($n = 7$). **Mean group change preclass to 6-month SAS TRAIT:** 41.7 to 37.6 in the intervention group ($n = 9$) vs 48.0 to 43.0 in the control group ($n = 3$). **Mean group change preclass to 6-week SDA:** 30.6 to 30.7 in the intervention group ($n = 19$) vs 30.9 to 29.1 in the control group ($n = 16$), *p*-value approaching significance. **Mean group change preclass to 6-month SDA:** 29.1 to 27.3 in the intervention group ($n = 13$) vs 29.0 to 29.4 in the control group ($n = 10$). **Mean group change preclass to 6-week SSS number:** 4.8 to 4.5 in the intervention group ($n = 19$) vs 3.3 to 2.9 in the control group ($n = 14$), $p < 0.05$. **Mean group change preclass to 6-month SSS number:** 4.5 to 3.7 in the intervention group ($n = 14$) ($p < 0.05$) vs 3.1 to 1.6 in the control group ($n = 10$) ($p < 0.05$). **Mean group change preclass to 6-week SSS satisfaction:** 5.6 to 5.6 in the intervention group ($n = 19$) vs 4.7 to 4.9 in the control group ($n = 14$). **Mean group change preclass to 6-month SSS satisfaction:** 5.6 to 5.0 in the intervention group ($n = 14$) ($p < 0.05$) vs 4.8 to 3.3 in the control group ($n = 10$) ($p < 0.05$). **Mean change preclass to 6-week BDI:** 1.7 in the intervention group ($n = 18$) vs −1.5 in the control group ($n = 16$). **Mean change preclass to 6-month BDI:** 0.9 in the intervention group ($n = 11$) vs −0.9 in the control group ($n = 10$). **Mean change 6-week to 6-month BDI:** 0 in the intervention group ($n = 11$) vs −0.7 in the control group ($n = 9$). **Mean change preclass to 6-week SAS STATS:** 3.5 in the intervention group ($n = 15$) vs −0.2 in the control group ($n = 10$). **Mean change preclass to 6-month SAS STATS:** 2.6 in the intervention group ($n = 11$) vs 3.0 in the control group ($n = 10$). **Mean change 6-week to 6-month SAS STATS:** −1.7 in the intervention group ($n = 11$) vs 3.8 in the control group ($n = 5$).

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			Mean change preclass to 6-week SAS TRAIT: 7.4 in the intervention group ($n = 14$) vs 1.4 in the control group ($n = 7$). **Mean change preclass to 6-week SDA**: –0.16 in the intervention group ($n = 19$) vs 1.8 in the control group ($n = 16$). **Mean change preclass to 6-month SDA**: 1.8 in the intervention group ($n = 13$) vs –0.4 in the control group ($n = 10$). **Mean change 6-week to 6-month SDA**: 1.2 in the intervention group ($n = 13$) vs –2.1 in the control group ($n = 9$). **Mean change preclass to 6-week SSS number**: 0.35 in the intervention group ($n = 19$) vs 0.4 in the control group ($n = 14$). **Mean change preclass to 6-month SSS number**: 0.7 in the intervention group ($n = 14$) vs 1.4 in the control group ($n = 10$). **Mean change 6-week to 6-month SSS number**: 0.8 in the intervention group ($n = 13$) vs 1.1 in the control group ($n = 9$). **Mean change preclass to 6-week SSS satisfaction**: 0 in the intervention group ($n = 19$) vs –0.12 in the control group ($n = 14$). **Mean change preclass to 6-month SSS satisfaction**: 0.67 in the intervention group ($n = 14$) vs 1.5 in the control group ($n = 10$). **Mean change 6-week to 6-month SSS satisfaction**: 0.6 in the intervention group ($n = 13$) vs 2.3 in the control group ($n = 9$) ($p < 0.05$)
Brugha et al., 2000,[196] UK Individual RCT 209 women	General hospital antenatal clinics At least 16 years, in first pregnancy and intending to continue to full term, within reasonable distance from hospital, could understand English, presence of any one of six depression items indicating depression on the modified GHQ	**Intervention:** Preparing For Parenthood (PFP) soon after the 28th week of gestation consisting of six structured 2-hour weekly antenatal classes, preceded by an initial introductory meeting with the participant and her partner. It ends with a postnatal reunion class when the babies are about 8 weeks old. **Enhancement:** Nurses and occupational therapists were trained to conduct PFP and attended supervision sessions **Usual care:** Standard antenatal care	**Threshold GHQ-D at 3 months:** 24 (26%) of the 94 women in the intervention group and 21 (22%) of the 96 women in the control group scored 2 or above on the GHQ-D, $OR_{adj} = 1.19$, 95% CI 0.59 to 2.37, $p = 0.62$, OR = 1.22, 95% CI 0.63 to 2.39, $p = 0.55$. **Threshold EPDS at 3 months:** 15 (16%) women in the intervention group and 18 (19%) women in the control group scored ≥ 11 on the EPDS, $OR_{adj} = 0.83$, 95% CI 0.39 to 1.79, $p = 0.64$, OR = 0.82, 95% CI 0.39 to 1.75, $p = 0.61$. SCAN ICD-10: 3 (3%) women in the intervention group and 6 (6%) women in the control group had ICD-10-diagnosed depression, $OR_{adj} = 0.47$, 95% CI 0.11 to 1.96, $p = 0.30$, OR = 0.48, 95% CI 0.12 to 1.99, $p = 0.30$. **Number of close relatives who had let her down:** 5 vs 6, OR = 0.84, 95% CI 0.25 to 2.86, $p = 0.78$. **Perceived poor social support at screening:** 6 vs 5, OR = 1.24, 95% CI 0.37 to 4.21, $p = 0.73$.

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Chen *et al.*, 2000,[214] Taiwan Individual RCT 115 women	Recruited from postnatal wards at two urban hospitals Inclusion criteria: over 18 years, survival of the infant, at least a junior high-school education and BDI = 9/10	**Intervention:** Four weekly support groups consisting of five to six mothers with their 6- to 10-week-old infants. A registered nurse researcher acted as the group leader **Usual care:** Not described – only states that the control group did not receive the support group intervention	**Any three poor social support variables:** 12 vs 10, OR = 1.26, 95% CI 0.52 to 3.07, $p = 0.61$. **High vs low confidence in ability to solve problems:** 7 vs 10, OR = 0.69, 95% CI 0.25 to 1.90, $p = 0.48$. **Avoidant vs approaching style of solving problems:** 45 vs 28, OR = 2.23, 95% CI 1.23 to 4.06, $p = 0.01$. **High vs low belief in personal control when solving problems:** 47 vs 45, OR = 1.13, 95% CI 0.64 to 2.00, $p = 0.67$. **High vs low belief in powerful others influencing life:** 10 vs 2, OR = 5.59, 95% CI 1.19 to 26.27, $p = 0.03$. **High vs low belief in chance or fate influencing their life:** 27 vs 27, OR = 1.03, 95% CI 0.55 to 1.93, $p = 0.93$. **High vs low belief in internal personal factors influencing their life:** 45 vs 47, OR = 0.96, 95% CI 0.54 to 1.69, $p = 0.88$. **Occurrence of many vs few non-pregnancy-specific events including life-threatening events:** 42 vs 36, OR = 1.35, 95% CI 0.75 to 2.40, $p = 0.31$. **Many vs few pregnancy vs specific events occurring:** 13 vs 22, OR = 0.54, 95% CI 0.25 to 1.15, $p = 0.11$. **Many vs few support-threatening events occurring:** 28 vs 30, OR = 0.93, 95% CI 0.50 to 1.73, $p = 0.83$. **Many vs few difficulties with activities of daily living:** 1 vs 6, OR = 0.16, 95% CI 0.02 to 1.37, $p = 0.09$. **High vs low dissatisfaction with housing:** 2 vs 4, OR = 0.50, 95% CI 0.09 to 2.80, $p = 0.43$. **Primary and secondary health service contact:** 13 vs 11, OR = 1.24, 95% CI 0.53 to 2.93, $p = 0.62$ **Within-group change mean BDI at week 4:** difference –6.60 (5.89) ($p < 0.01$) in the intervention group ($n = 30$) vs difference –1.40 (8.33) in the control group ($n = 30$). **Within-group change mean PSS at week 4:** difference –3.10 (4.53) ($p < 0.01$) in the intervention group ($n = 30$) vs difference –1.30 (4.26) in the control group ($n = 30$). **Within-group change mean ISEL at week 4:** difference 2.60 (5.08) ($p < 0.01$) in the intervention group ($n = 30$) vs difference 0 (5.14) in the control group ($n = 30$). **Within-group change mean SEI at week 4:** difference 1.03 (4.57) in the intervention group ($n = 30$) vs difference 1.03 (2.97) in the control group ($n = 30$). **Threshold BDI at week 4:** 10 (33.3%) of the 30 women in the intervention group and 18 (60.0%) of the 30 women in the control group had BDI scores of ≥ 10 ($p < 0.05$)

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Elliot, 2000,[199] UK Individual RCT (allocated based on expected date of delivery) 99 women	Hospital antenatal clinic Women were classified as vulnerable if they scored 2 on any one of the vulnerability questions in the Leverton questionnaire, or scored 1 on more than one question, or scored ≥ 10 on the Crown–Crisp Experimental Index anxiety subscale. In addition, second-time mothers were asked whether they felt more tense or depressed than usual after the birth of their child	**Intervention:** Combined psychoeducation and social support in a programme with continuity of care across the childbirth transition. Women were invited to take part in Preparation for Parenthood (for first-time mothers) or Surviving Parenthood (for second-time mothers). Health visitors were asked to make a mid-pregnancy visit. Group intervention consisting of five monthly meetings scheduled antenatally, starting around 24 weeks of pregnancy, and six monthly meetings scheduled postnatally **Usual care:** Not described	**Bedford College criteria in first 3 months:** significant difference was obtained in the number of borderline or case depressions in the first 3 months on the Bedford College criteria (chi-squared, *p* < 0.05); **Catego classification:** similar although not quite significant, results were obtained with the Catego classification when other diagnoses were excluded (chi-squared, *p* < 0.1). **Bedford College criteria:** depression rate in those invited was 19% (9 out of 47), half that in the not invited (39%, 20 out of 52). Further outcome data were reported for first-time and second-time mothers separately but are not reported here
O'Hara et al., 2000,[213] USA Individual RCT 120 women	Unclear Women delivering in four counties who were over 18 years, were married or living with a partner for at least 6 months, met criteria for depression on the Inventory to Diagnose Depression and DSM-IV (SCID) and had a minimum score of 12 on the amended 17-item version of the HRSD	**Intervention:** 12 × 1- hour weekly sessions of interpersonal psychotherapy (IPT). Enhancement: Therapist read and became familiar with the *Interpersonal Psychotherapy of Depression* manual and the manual of *Interpersonal Psychotherapy for Postpartum Depression*. Therapists received 40 hours of didactic lectures and videotape presentations, meeting the standard training of IPT therapists used in extramural research projects. They were also required to complete a 12-session course of IPT with a postnatally depressed woman at a satisfactory level of competence and adherence before entering the treatment phase of the study. Therapists were continually monitored during the trial for adherence to the IPT treatment manuals **Usual care:** Women in the control group waited 12 weeks before receiving the treatment. Although no therapy was provided during this time, clinical assessments using the HRSD were conducted at 4, 8 and 12 weeks after allocation. Brief telephone contacts were made at 2, 6 and 10 weeks to evaluate suicide risk and ability to wait for the treatment	**Mean HRSD at 4 weeks:** 19.4 (4.6) in the intervention group (*n* = 48) vs 19.8 (5.3) in the control group (*n* = 51), *p* = 0.007. **Mean HRSD at 8 weeks:** 12.6 (7.0) in the intervention group (*n* = 48) vs 16.4 (6.5) in the control group (*n* = 51), *p* = 0.006. **Mean HRSD at 12 weeks:** 8.3 (5.3) in the intervention group (*n* = 48) vs 16.8 (8.4) in the control group (*n* = 51), *p* < 0.001. **Mean BDI at 4 weeks:** 17.7 (8.0) in the intervention group (*n* = 48) vs 21.6(8.1) in the control group (*n* = 51), *p* = 0.02. **Mean BDI at 8 weeks:** 13.6 (7.5) in the intervention group (*n* = 48) vs 19.1 (8.9) in the control group (*n* = 51), *p* = 0.001. **Mean BDI at 12 weeks:** 10.6 (6.8) in the intervention group (*n* = 48) vs 19.2 (8.7) in the control group (*n* = 51), *p* < 0.001. **Group × assessment occasion interaction:** ITT HRSD scores: in favour of IPT, *p* = 0.003; completer HRSD scores: *p* < 0.001; ITT BDI scores: *p* < 0.001; completer BDI scores: *p* < 0.001. **Threshold HRSD:** ITT: 41 (68.3%) of the 60 women in the intervention group vs 51 (85.0%) of the 60 women in the control group had HRSD ≥ 7, *p* = 0.03; completers: 31 (64.6%) of the 48 women in the intervention group vs 44 (86.3%) of the 51 women in the control group had HRSD ≥ 7, *p* = 0.007.

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			Threshold BDI: ITT: 37 (61.7%) of the 60 women in the intervention group vs 49 (81.7%) of the 60 women in the control group had BDI ≥ 10, $p = 0.02$; completers: 27 (56.2%) of the 48 women in the intervention group vs 44 (86.3%) of the 51 women in the control group had BDI ≥ 10, $p = 0.001$. **DSM-IV major depression:** completers: 6 (12.5%) of the 48 women in the intervention group vs 35 (68.6%) of the 51 women in the control group met DSM-IV criteria, $p < 0.001$. **Reduction of HRSD scores ≥ 50%:** completers: 30 (62.5%) of the 48 women in the intervention group vs 9 (17.6%) of the 51 women in the control group had reduction of HRSD ≥ 50%, $p < 0.001$. **Reduction of BDI scores ≥ 50%:** completers: 29 (60.4%) of the 48 women in the intervention group vs 8 (15.7%) of the 51 women in the control group had reduction of BDI ≥ 50%, $p < 0.001$. **Mean SAS-SR at 4 weeks:** 2.26 (0.35) in the intervention group ($n = 48$) vs 2.47 (0.40) in the control group ($n = 51$), $p = 0.008$. **Mean SAS-SR at 8 weeks:** 2.05 (0.33) in the intervention group ($n = 48$) vs 2.36 (0.42) in the control group ($n = 51$), $p < 0.001$. **Mean SAS-SR at 12 weeks:** 1.93 (0.34) in the intervention group ($n = 48$) vs 2.35 (0.45) in the control group ($n = 51$), $p < 0.001$. **Mean PAS at 4 weeks:** 2.59 (0.36) in the intervention group ($n = 48$) vs 2.66 (0.32) in the control group ($n = 51$), $p = $ not significant. **Mean PAS at 8 weeks:** 2.44 (0.31) in the intervention group ($n = 48$) vs 2.62 (0.36) in the control group ($n = 51$), $p = 0.01$. **Mean PAS at 12 weeks:** 2.33 (0.29) in the intervention group ($n = 48$) vs 2.57 (0.38) in the control group ($n = 51$), $p = 0.001$. **Mean DAS at 4 weeks:** 97.3 (20.3) in the intervention group ($n = 48$) vs 87.0 (25.0) in the control group ($n = 51$), $p = 0.03$. **Mean DAS at 8 weeks:** 100.4 (19.3) in the intervention group ($n = 48$) vs 88.7 (25.6) in the control group ($n = 51$), $p = 0.01$. **Mean DAS at 12 weeks:** 101.2 (20.7) in the intervention group ($n = 48$) vs 88.7 (27.6) in the control group ($n = 51$), $p = 0.01$

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Gorman, 2001,[200] USA Individual RCT 45 high-risk women	Recruited pregnant women in their third trimester attending prenatal appointment at obstetric clinics Women with history of a major depressive disorder, a score > 12 on the BDI, first-degree relative who required psychiatric treatment, DAS score < 100 or experiencing two or more life events predicted to have a moderate to severe impact	**Intervention:** Preventive intervention incorporating principles of IPT (P-IPT). The first two sessions took place during pregnancy between 32 weeks' gestation and delivery. The remaining three sessions occurred during the postnatal period, beginning at the second week postnatally. A treatment manual describing the protocol intervention provided guidance for individual session formats and goals **Usual care:** No treatment control group	**Major depression at 1 month:** 0/20 women in the intervention group and 5/20 (25%) women in the control group met DSM-III-R criteria, $p = 0.02$. **Major depression at 6 months:** 3/20 (15%) women in the intervention group and 4/19 (22%) women in the control group met DSM-III-R criteria, $p = 0.47$. **Mean BDI 1 month:** 9.1 (6.7) intervention ($n = 17$) vs 11.3 (6.0) control ($n = 15$). **Mean BDI 6 months:** 10.7 (9.9) intervention ($n = 13$) vs 11.3 (7.8) control ($n = 17$). **Mean EPDS 1 month:** 7.2 (5.3) intervention ($n = 18$) vs 7.5 (4.2) control ($n = 15$). **Mean EPDS 6 months:** 7.9 (5.2) intervention ($n = 13$) vs 8.0 (5.6) control ($n = 17$). **Mean PANAS(−) 1 month:** 17.9 (6.7) intervention ($n = 17$) vs 19.5 (4.6) control ($n = 17$). **Mean PANAS(−) 6 months:** 20.5 (7.8) intervention ($n = 13$) vs 21.2 (7.8) control ($n = 16$). **Mean PANAS(+) 1 month:** 30.0 (7.7) intervention ($n = 17$) vs 29.4 (8.2) control ($n = 17$). **Mean PANAS(+) 6 months:** 30.4 (8.4) intervention ($n = 13$) vs 28.4 (6.5) control ($n = 16$). **Mean SCL-90-R depression 1 month:** 0.71 (0.63) intervention ($n = 17$) vs 1.00 (0.56) control ($n = 17$). **Mean SCL-90-R depression 6 months:** 1.12 (0.82) intervention ($n = 13$) vs 1.09 (0.84) control ($n = 17$)
Horowitz et al., 2001,[221] USA Individual RCT 122 women	Women were recruited from one large teaching hospital and another hospital. Data were collected over the phone and/or from home EPDS ≥ 10, women with healthy babies (defined as those infants discharged with a normal newborn examination) and who could speak English	**Intervention:** Women received three home visits when their babies were 4- to 8-weeks-old (time 1), 10- to 14-weeks-old (time 2) and 14- to 18-weeks-old (time 3). In addition, women received a coached behavioural intervention designed to promote maternal–infant responsiveness. **Enhancement:** Intensive training in all study procedures and assignment of roles for research nurses **Usual care:** Women received three home visits when their babies were 4- to 8-weeks-old (time 1), 10- to 14-weeks-old (time 2) and 14- to 18-weeks-old (time 3).	**Mean Dyadic Mutuality Code (DMC) level of responsiveness in maternal–infant relationship:** time 1: 8.83 (1.76) in the intervention group ($n = 60$) vs 8.67 (1.64) in the control group ($n = 57$); time 2: 9.73 (1.65) in the intervention group vs 8.77 (1.72) in the control group, $p = 0.02$; time 3: 9.55 (1.77) in the intervention group vs 8.80 (1.86) in the control group, $p = 0.03$. **Repeated measures ANOVA DMC:** between subjects $p = 0.006$; within subjects time $p = 0.025$, time by group $p = 0.12$. **Mean BDI: time 1:** 15.5 (1.17), time 2: 10.99 (0.96), time 3: 10.27 (0.99) in the intervention group ($n = 60$) vs time 1: 13.24 (0.92), time 2: 10.10 (0.84), time 3: 9.51 (0.77) in the control group ($n = 57$). **Repeated measures ANOVA BDI:** between subjects $p = 0.67$, within subjects time $p < 0.0005$, time by group $p = 0.67$

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Prendergast and Austin, 2001,[212] Australia Individual RCT 37 women	Unclear Women attending regular screening by the early childhood nurses (ECNs) in the postnatal period with EPDS ≥ 12 and DSM-IV criteria for major or minor depression	**Intervention:** 6 × 1-hour home-based cognitive behavioural therapy (CBT) sessions. Enhancement: Five self-selected ECNs were trained by a psychiatrist, psychologist and senior psychiatry registrar in the modified CBT method. The basis of the training was a CBT-based workbook prepared specifically for the study. The ECNs received weekly group supervision **Usual care:** Six weekly clinic visits for ideal standard care	**Mean EPDS at baseline:** 15.9 (2.8) in the intervention group (*n* = 17) vs 13.7 (2.3) in the control group (*n* = 20), *p* = 0.03. **Mean EPDS week 6:** 8.1 (2.9) in the intervention group (*n* = 17) vs 6.5 (6.2) in the control group (*n* = 20), *p* = not significant. **Mean EPDS week 24:** 6.2 (4.2) in the intervention group (*n* = 15) vs 7.7 (3.9) in the control group (*n* = 18), *p* = not significant. **Mean MADRS baseline:** 21.7 (3.6) in the intervention group (*n* = 17) vs 20.0 (5.0) in the control group (*n* = 20), *p* = not significant. **Mean MADRS week 6:** 8.4 (5.3) in the intervention group (*n* = 17) vs 12.1 (8.3) in the control group (*n* = 20), *p* = not significant. **Threshold EPDS at week 6:** 3 (18%) of the 17 women in the intervention group vs 5 (23%) of the 20 women in the control group had EPDS ≥ 10. **Threshold EPDS at week 24:** 1 (7%) of the 15 women in the intervention group vs 3 (17%) of the 18 women in the control group had EPDS ≥ 10
Zlotnick *et al.*, 2001,[204] USA Individual RCT 25 women	Attending prenatal clinic at general hospital Women receiving public assistance who were between 20 and 32 weeks' gestation and who reported at least one predictor of PND (previous episode of depression or PND, mild to moderate levels of depressive symptoms, poor social support or life stressor within last 6 months)	**Intervention:** Survival skills for new mothers, involving 4 × 60-minute group sessions over a 4-week period **Usual care:** Treatment as usual	**BDI change score 3 months:** greater change in BDI scores in the 17 women in the intervention group than in the 18 in the control group, *p* = 0.001. **Mean BDI:** 8.4 (7.8) intervention (*n* = 17) vs 11.3 (4.8) control (*n* = 18). **Reliable improvement after intervention:** 6 (35%) of the 17 women in the intervention group and 2 (11%) of the 18 women in the control group showed reliable improvement. **DSM-IV major depression:** 0 (0%) of the 17 women in the intervention group and 6 (33%) of the 18 women in the control met DSM-IV criteria for major depression
Chabrol *et al.*, 2002,[215] France Individual RCT 258 women	Three obstetric clinics French-speaking women without current treatment with a psychiatrist or a psychologist and with EPDS ≥ 9	**Intervention:** Received a 1-hour prevention session between 2 and 5 days postnatally during their stay at the clinic. The prevention session comprised: (1) educational information about the realities of parenthood and guidance for normal infant development problems; (2) supportive component; (3) cognitive behavioural component. Subjects with EPDS ≥ 11 (postal questionnaire) at 4–6 weeks postnatally were interviewed at home and those in the intervention group were invited to participate in 1-hour weekly home visits for 5–8 weeks. The home-based treatment integrated supportive, educational, cognitive behavioural and psychodynamic components. **Enhancement:** Therapists who had implemented the prevention session already participated in didactic and clinical training including discussions and role-playing exercises.	

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
		Weekly supervision was provided to ensure therapists adherence to the treatment guide **Usual care:** In the first part of the study control women were given a telephone number to enable them to contact the student who they met at the clinic. In the second part depressed women were called each week for a brief assessment by the student to ensure that they could wait for treatment.	**Prevention intervention: Threshold EPDS at 4–6 weeks:** 29 (30.2%) women in the intervention group and 55 (48.2%) women in the control group had EPDS scores of ≥ 11 ($p = 0.0067$). **Group by assessment occasion:** Repeated measures ANOVA revealed a statistically significant group × assessment occasion interaction [$F_{(1,212)} = 30.9$, $p < 0.0001$]. **Mean EPDS scores at 4–6 weeks:** 8.5 (4.0) in the intervention group ($n = 97$) vs 10.3 (4.4) in the control group ($n = 114$), $p = 0.002$. ITT analyses were performed on all subjects with EPDS ≥ 9 in the first days postnatally. LOCF was used for women with missing EPDS scores at time 2. A significant group × assessment interaction in favour of the intervention group was found ($p < 0.01$) **Treatment intervention: Mean HDRS at 10–12 weeks:** 5.7 (3.3) in the intervention group ($n = 18$) vs 16.2 (4.5) in the control group, $p < 0.0001$. **Mean BDI at 10–12 weeks:** 4.7 (3.0) in the intervention group ($n = 18$) vs 15.7 (4.4) in the control group, $p < 0.0001$. **Mean EPDS at 10–12 weeks:** 5.9 (2.7) in the intervention group ($n = 18$) vs 13.7 (3.6) in the control group, $p < 0.0001$; **Group by assessment occasion:** MANOVA detected a significant group × assessment occasion interaction in favour of the intervention ($p < 0.0001$); repeated measures ANOVA showed a significant group × assessment interaction for each treatment outcome: HDRS ($p < 0.0001$), BDI ($p < 0.0001$) and EPDS ($p < 0.0001$). **Threshold HDRS:** 12 (66.6%) of the 18 women in the intervention group vs 2 (6.6%) of the 30 women in the control group had HDRS < 7, $p < 0.0001$. **Threshold BDI:** 11 (61.1%) of the 18 women in the intervention group vs 1 (3.3%) of the 30 women in the control group had BDI < 4, $p < 0.0001$. Reduction of HDRS scores ≥ 50%: 15 (83.3%) of the 18 women in the intervention group vs 4 (13.3%) of the 30 women in the control group, $p < 0.0001$. Reduction of BDI scores ≥ 50%: 16 (88.8%) of the 18 women in the intervention group vs 4 (13.3%) of the 30 women in the control group, $p < 0.0001$. Reduction of EPDS scores ≥ 50%: 13 (72.2%) of the 18 women in the intervention group vs 1 (3.3%) of the 30 women in the control group, $p < 0.0001$.

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			ITT analyses were also performed on all subjects with EPDS $\geq$ 11 during 4–6 weeks postnatally. LOCF was used for women with missing EPDS scores. A significant group × assessment interaction in favour of the intervention group was found [$F_{(1,84)} = 32.1$, $p < 0.0001$]
Honey *et al.*, 2002,[209] UK Individual RCT 45 women	Baby clinics for recruitment. Assessment in the home Attended mother and baby clinics in Gwent, EPDS > 12, not exhibiting psychotic symptoms and their most recent child was < 12 months previously	**Intervention:** Brief psychoeducational intervention consisting of eight times weekly 2-hour meetings run by two health visitors, which had three components: (1) educational – providing information on PND, strategies for coping with difficult childcare situations and eliciting social support; (2) use of cognitive behavioural techniques to tackle women's erroneous cognitions about motherhood and to provide strategies for coping with anxiety; and (3) teaching the use of relaxation **Usual care:** Not described	**Mean EPDS week 8:** 14.87 (5.97) in the intervention group (*n* = 23) vs 16.95 (5.44) in the control group (*n* = 22). **Mean EPDS week 24:** 12.55 (4.62) in the intervention group (*n* = 23) vs 15.63 (7.28) in the control group (*n* = 22). **ANOVA** with group (intervention and control) and time as factors revealed an effect of group (*p* = 0.01) and time (*p* < 0.001) and a significant interaction between group and time (*p* < 0.05). **Threshold EPDS at week 8:** 15 (65%) of the 23 women in the intervention group vs 16 (64%) of the 22 women in the control group had EPDS $\geq$ 13. **Threshold EPDS at week 24:** 7 (35%) of the 20 women in the intervention group vs 12 (64%) of the 19 women in the control group had EPDS $\geq$ 13. **Additional outcomes:** There were no effects of time, group or interactions between time and group on total social support [maximum $F_{(1,43)} = 2.29$, $p > 0.1$], marital relationship [maximum $F_{(1,34)} < 1$, $p > 0.1$] or coping scales [maximum $F_{(1,32)} = 4.04$, $p > 0.05$ but $p < 0.1$]
Armstrong and Edwards, 2003,[206] Australia Individual RCT 20 women	Women recruited by recommendations from various health professionals and via advertising Women who had a child, between 6 weeks and 12 months postpartum, EPDS $\geq$ 12 and passed fitness test	**Intervention:** 12-week multi-intervention programme. Exercise component involved pram walking three times per week with the group for 30–40 minutes at a moderate intensity (60–75% of age-predicted heart rate). Support sessions were also encouraged after one of the weekly pram-walking sessions **Usual care:** Phone support was provided to participants at week 6 and they were encouraged to contact researchers if they had any concerns. Women were also advised to maintain their usual regime and social activities	**Mean EPDS at week 6:** 7.20 (4.32) in the intervention group (*n* = 10) vs 13.50 (4.53) in the control group (*n* = 10). **Mean EPDS at week 12:** 4.60 (3.34) in the intervention group (*n* = 10) vs 14.70 (7.66) in the control group (*n* = 10). **Mean DASS at week 6:** 0.08 (0.92) in the intervention group (*n* = 10) vs 1.80 (1.69) in the control group (*n* = 10). **Mean DASS at week 12:** 0 (0) in the intervention group (*n* = 10) vs 1.70 (1.57) in the control group (*n* = 10). **Mean GHQ at week 6:** 2.44 (1.81) in the intervention group (*n* = 10) vs 5.70 (4.00) in the control group (*n* = 10). **Mean GHQ at week 12:** 1.11 (1.17) in the intervention group (*n* = 10) vs 5.10 (4.09) in the control group (*n* = 10). **Mean SSI at week 6:** 101.60 (20.71) in the intervention group (*n* = 10) vs 84.10 (19.65) in the control group (*n* = 10). **Mean SSI at week 12:** 101.60 (19.31) in the intervention group (*n* = 10) vs 89.00 (17.36) in the control group (*n* = 10). **Mean volume of oxygen scores at week 12:** 28.53 ml/kg/min in the intervention group vs 19.35 m/l/kg/min in the control group

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Cooper et al., 2003,[216] Murray et al., 2003,[217] UK Individual RCT 193 women	Maternity hospital Primiparous, living within a 15-mile radius of the maternity hospital, English as first language, EPDS ≥ 12 and DSM-III-R major depressive disorder	**Intervention:** Three treatments all focusing (in different ways) on the mother–infant relationship: CBT, psychodynamic therapy (PT) and non-directive counselling (NDC). Therapy was conducted in the women's own homes on a weekly basis from 8 weeks to 18 weeks postnatally. **Enhancement:** Training in CBT and counselling treatments **Usual care:** Routine primary care involving the normal care provided by the primary health-care team (i.e. general practitioners and health visitors) with no additional input (apart from assessment) from the research team	**Mean EPDS at 4.5 months:** 9.4 (5.0) across all treatments (n = 134) [9.9 (5.9) in the NDC group (n = 47), 8.9 (4.2) in the PT group (n = 45), 9.2 (4.8) in the CBT group (n = 42)] vs 11.3 (4.8) in the control group (n = 50), mean difference –1.9, 95% CI –3.5 to –0.3. **Mean EPDS at 4.5 months adjusted for mean centred baseline EPDS score:** all treatments vs control mean difference –2.5, 95% CI –3.9 to –1.0, p ≤ 0.001; NDC vs control mean difference –2.1, 95% CI –3.8 to –0.3, p = 0.02; PT vs control mean difference –2.6, 95% CI –4.4 to –0.9, p = 0.003; CBT vs control mean difference –2.7, 95% CI –4.5 to –0.9, p = 0.003. **Mean EPDS at 9 months:** 9.3 (5.5) across all treatments (n = 129) [9.6 (5.8) in the NDC group (n = 46), 9.5 (5.5) in the PT group (n = 43), 8.6 (5.9) in the CBT group (n = 40)] vs 9.2 (5.4) in the control group (n = 48), mean difference 0.1, 95% CI –1.7 to 1.9. **Mean EPDS at 9 months adjusted for mean centred baseline EPDS score:** all treatments vs control mean difference –0.3, 95% CI –2.0 to 1.3, p = 0.70; NDC vs control mean difference –0.2, 95% CI –3.5 to 3.2, p = 0.87; PT vs control mean difference 0.2, 95% CI –2.9 to 3.3, p = 0.85; CBT vs control mean difference –1.0, 95% CI –4.4 to 2.4, p = 0.33. **DSM-III-R diagnosis at 4.5 months:** 53 (39%) of the 135 women across all treatments [22 (46%) of the women in the NDC group (n = 48), 13 (29%) of the women in the PT group (n = 45), 18 (43%) of the women in the CBT group (n = 42)] vs 30 (60%) of the 50 women in the control group had DSM-III-R diagnosis. **DSM-III-R diagnosis at 4.5 months adjusted for mean centred baseline SCID score:** all treatments vs control RR 1.60, 95% CI 1.14 to 1.98, p = 0.01; NDC vs control RR 1.38, 95% CI 0.82 to 1.89, p = 0.14; PT vs control RR 1.89, 95% CI 1.33 to 2.23, p = 0.002; CBT vs control RR 1.50, 95% CI 0.92 to 1.98, p = 0.09. **DSM-III-R diagnosis at 9 months:** 35 (27%) of the 130 women across all treatments [16 (34%) of the women in the NDC group (n = 47), 9 (21%) of the 43 women in the PT group, 10 (25%) of the women in the CBT group (n = 40)] vs 15 (31%) of the 48 women in the control group had DSM-III-R diagnosis.

continued

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			DSM-III-R diagnosis at 9 months adjusted for mean centred baseline SCID score: all treatments vs control RR 1.09, 95% CI 0.83 to 1.26, $p = 0.48$; NDC vs control RR 0.99, 95% CI 0.33 to 1.36, $p = 0.77$; PT vs control RR 1.15, 95% CI 0.54 to 1.39, $p = 0.28$; CBT vs control RR 1.12, 95% CI 0.45 to 1.39, $p = 0.56$. **Expertise effects specialist vs non-specialist controlling for type of treatment, baseline EPDS, social adversity and level of education:** non-specialists vs specialists at 4.5 months mean difference in EPDS –2.1, 95% CI –3.7 to –0.5, $p = 0.01$ and at 9 months –2.0, 95% CI –4.0 to –0.1, $p = 0.04$; non-specialist NDC –2.2, 95% CI –4.0 to –0.3, $p = 0.02$; non-specialist PT –3.3, 95% CI –5.4 to –1.1, $p = 0.003$; non-specialist CBT –2.4, 95% CI –4.1 to –0.8, $p = 0.003$; at 9 months there were no statistically significant differences between the control group and each of the non-specialist groups. **Behavioural management problems:** 9 (47%) of the 19 women receiving NDC vs 8 (44%) of the 18 women in the control group still reported problems after the intervention, % difference –3, 95% CI –35 to 29; 15 (68%) of the 22 women receiving PT vs 8 (44%) of the 18 women in the control group still reported problems after the intervention, % difference –24, 95% CI –54 to 6; 9 (41%) of the 22 women receiving CBT vs 8 (44%) of the 18 women in the control group still reported problems after the intervention, % difference 3, 95% CI –28 to 34. **Relationship problems:** 18 (72%) of the 25 women receiving NDC vs 19 (83%) of the 23 women in the control group still reported problems after the intervention, % difference 11, 95% CI –12 to 34; 12 (50%) of the 24 women receiving PT vs 19 (83%) of the 23 women in the control group still reported problems after the intervention, % difference 33, 95% CI 8 to 58; 12 (41%) of the 29 women receiving CBT vs 19 (83%) of the 23 women in the control group still reported problems after the intervention, % difference 42, 95% CI –18 to 66. **Behavioural management problems at 4.5 months adjusted:** NDC vs control RR 0.91, 95% 0.42 to 1.58, $p = 0.77$; PT vs control RR 1.21, 95% CI 0.62 to 1.87, $p = 0.52$; CBT vs control RR 0.83, 95% CI 0.37 to 1.50, $p = 0.60$. **Relationship problems at 4.5 months adjusted:** NDC vs control RR 0.63, 95% 0.32 to 0.97, $p = 0.03$; PT vs control RR 0.57, 95% CI 0.28 to 0.92, $p = 0.01$; CBT vs control RR 0.46, 95% CI 0.20 to 0.81, $p = 0.002$.

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			Mother–infant interactions at 4.5 months: the three treatments were found to be comparable with the control after adjusting for the baseline measure. Therapist effects at 4.5 months: non-specialists vs specialists behavioural problems RR 1.34, 95% CI 0.80 to 1.91, *p* = 0.24; non-specialists vs specialists relationship problems RR 1.23, 95% CI 0.76 to 1.68, *p* = 0.36; non-specialists vs specialists maternal sensitivity –0.18, 95% CI –0.42 to 0.07, *p* = 0.15
Dennis, 2003,[207] Canada	Immunisation clinics managed by public health nurses	**Intervention:** Access to standard community postnatal services in addition to telephone-based peer support initiated within 48–72 hours of randomisation. Women who had a	**Threshold EPDS at 4 weeks:** 9 (45%) of the 20 women in the intervention group vs 16 (73%) of the 22 women in the control group had EPDS > 9, OR = 3.26, 95% CI 0.90 to 11.81, *p* = 0.06; 2 (10%) of the 20 women in the
Individual RCT	New mothers between 8 and 12 weeks postnatally, at	history of and recovery from PND with a desire to help new mothers volunteered to give the telephone-based support.	intervention group vs 9 (41%) of the 22 women in the control group had EPDS > 12, OR = 6.23, 95% CI 1.15 to 33.77, *p* = 0.02. **Threshold EPDS at 8 weeks:** 7 (35%) of
42 women	least 18 years, were able to speak English, had a singleton	**Enhancement:** Volunteers received a 4-hour training session	the 20 women in the intervention group vs 16 (76%) of the 21 women in the control group had EPDS > 9, OR = 5.94,
	birth at 37 weeks' gestation or more, EPDS > 9, resided	**Usual care:** Standard community postnatal services	95% CI 1.52 to 23.18, *p* = 0.008; 3 (15%) of the 20 women in the intervention group vs 11 (52%) of the 21
	in surrounding region, were accessible by local telephone		women in the control group had EPDS > 12, OR = 6.23, 95% CI 1.40 to 27.84, *p* = 0.01. **Mean EPDS at 8 weeks:**
	call, were not currently taking antidepressants, did not have a		after the baseline characteristics were controlled for the results indicate that the peer support intervention
	history of psychotherapy during last 12 months and no history of		significantly decreased EPDS scores at the 8-week assessment, OR = 4.7, 95% CI 0.91 to 25.46. **Mean EPDS**
	chronic depression, psychiatric clinical disorder or postnatal		**at 4 weeks:** 8.5 (3.7) in the intervention group (*n* = 20) vs 12.1 (4.6) in the control group (*n* = 22), *p* = 0.008. **Mean**
	psychosis		**EPDS at 8 weeks:** similar group differences were found (*p* = 0.006). **Maternal self-esteem 8 weeks:** 30.00 (4.21) in the intervention group (*n* = 20) vs 30.00 (4.21) in the
			control group (*n* = 22). **Childcare stress 8 weeks:** 4.95 (2.68) in the intervention group (*n* = 20) vs 6.48 (3.63) in the control group (*n* = 21). **Maternal loneliness 8 weeks:**
			20.37 (5.23) in the intervention group (*n* = 20) vs 23.91 (6.07) in the control group (*n* = 21). Further outcome data were reported regarding perceptions of peer support but are not reported here

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Heh and Fu, 2003,[208] Taiwan Individual RCT 70 women	Two general hospitals Married women, first-time mothers aged between 20 and 35 years, normal spontaneous delivery with single full-term healthy baby (gestation 38–42 weeks, body weight > 2500 g and APGAR score > 8). The EPDS was sent to 500 women in the fourth week after giving birth; 407 returned the questionnaire and those scoring ≥ 10 were invited to take part in the study	**Intervention:** Informational support booklet (three-page booklet developed specifically for the study) on postnatal depression 6 weeks postnatally **Usual care:** Not described. The informational support booklet was also provided to the women in the control group after assessing their depression scores at 3 months postnatally	**Threshold EPDS at 3 months:** 14 (40%) of the 35 women in the intervention group vs 24 (69%) of the 35 in the control group had EPDS ≥ 10, $p = 0.02$. **Within intervention group mean EPDS change:** 16.5 (3.0) at 4 weeks ($n = 35$) vs 10.8 (4.4) at 3 months ($n = 35$), $p < 0.0005$. **Within control group mean EPDS change:** 16.3 (2.7) at 4 weeks ($n = 35$) vs 12.1 (3.0) at 3 months ($n = 35$), $p < 0.0005$. **Mean EPDS at 3 months:** 10.8 (4.4) in the intervention group ($n = 35$) vs 12.1 (3.0) in the control group ($n = 35$), $p = 0.02$. Further outcome data were reported regarding perceptions of the intervention but are not reported here
Misri et al., 2004,[211] Canada Individual RCT 35 women	Women recruited from outpatient referrals to the Reproductive Mental Health Programme at a tertiary care hospital Women were included if between 18 and 40 years, had delivered a healthy baby close to term with a minimum birthweight of 2.5 kg, were non-smokers, met DSM-IV criteria for PND or anxiety disorders, had EPDS ≥ 12, HAMD 18 and HAMA ≥ 20	**Intervention:** Prescribed paroxetine, with the initial dose at 10 mg. An individually tailored, flexible-dose regimen was available up to a maximum of 50 mg daily. In addition, women received a 1-hour individual CBT session every week for 12 weeks **Usual care:** Same protocol for paroxetine as the intervention group	**Mean HAMD at 3 months:** 6.00 (7.09) in the intervention group ($n = 19$) vs 4.50 (4.27) in the control group ($n = 16$), $p = 0.47$. **Mean HAMA at 3 months:** 6.68 (7.54) in the intervention group ($n = 19$) vs 6.06 (5.92) in the control group ($n = 16$), $p = 0.79$. **Mean EPDS at 3 months:** 8.64 (5.51) in the intervention group ($n = 14$) vs 9.71 (5.17) in the control group ($n = 14$), $p = 0.60$. **Mean Yale–Brown Obsessive Compulsive Scale (YBOCS) at 3 months:** 3.26 (5.91) in the intervention group ($n = 19$) vs 2.13 (4.92) in the control group ($n = 16$), $p = 0.55$. **Mean clinical global impression scale at 3 months:** 1.78 (1.22) in the intervention group ($n = 19$) vs 1.44 (0.63) in the control group ($n = 16$), $p = 0.32$. **HAMD responder (≥ 50% score reduction):** 15 (78.9%) of the 19 women in the intervention group vs 14 (87.5%) of the 16 women in the control group, $p = 0.5$. **HAMA responder (≥ 50% score reduction):** 16 (84.2%) of the 19 women in the intervention group vs 12 (75.0%) of the 16 women in the control group, $p = 0.5$. **EPDS responder (≥ 50% score reduction):** 7 (58.3%) of the 12 women in the intervention group vs 8 (61.5%) of the 13 women in the control group, $p = 0.87$. **YBOCS responder (≥ 60% score reduction):** 11 (78.6%) of the 14 women in the intervention group vs 8 (80.0%) of the 10 women in the control group, $p = 0.93$. **Depression (based on HAMD):** 12 (63.2%) of the 19 women in the intervention group and 12 (75.0) of the 16 women in the control group, $p = 0.23$.

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			Anxiety (based on HAMA): 11 (57.9%) of the 19 women in the intervention group and 12 (75.0) of the 16 women in the control group, *p* = 0.45. **Obsessions and/or OCD (based on YBOCS):** 10 (71.4%) of the 14 women in the intervention group vs 8 (80%) of the 10 women in the control group, *p* = 0.50. **Mean week of recovery:** 9.50 weeks in the intervention group vs 11.16 weeks in the control group. **Mean dose:** 32.50mg in the intervention group vs 36.25mg in the control group
Zayas *et al.*, 2004,[202] McKee *et al.*, 2006,[201] USA Individual RCT 100 women	Three community health centres Low-income, ethnic minority (Hispanic or African American) women 18 years or older receiving prenatal care for low-risk pregnancies who had not reached their 32nd week of gestation, did not have a major mental illness or significant medical or obstetric complication, including HIV, with BDI-II ≥ 14	**Intervention:** Multicomponent psychosocial intervention: eight session of CBT, four child development psychoeducational modules, social support building sessions twice a month. The sessions took place in women's homes or at the health centres. Trained and supervised in the intervention **Usual care:** Usual clinic services, which consisted of typical psychosocial services not using specific methods. Offered counselling, support groups, psychoeducational workshops and family-orientated interventions	**Mean BDI-II in third trimester:** 20.5 (6.2) in the intervention group (*n* = 57) vs 22.4 (7.0) in the control group (*n* = 43). **Mean BDI-II at 3 months:** 12.1 (6.4) in the intervention group (*n* = 57) vs 12.6 (7.0) in the control group (*n* = 43). **Mean total functional support scores NSSQ TFS in third trimester:** 97.4 (68.5) in the intervention group (*n* = 57) vs 75.7 (46) in the control group (*n* = 43). **Mean total functional support scores NSSQ TFS at 3 months:** 84.3 (44.1) in the intervention group (*n* = 57) vs 84.6 (41.4) in the control group (*n* = 43). **Mean mother–infant interactions (mother ratings):** state 2.86 (0.36) in the intervention group (*n* = 21) vs 2.75 (0.44) in the control group (*n* = 20); activity 2.48 (0.60) in the intervention group (*n* = 21) vs 2.35 (0.81) in the control group (*n* = 20); orientation 2.67 (0.58) in the intervention group (*n* = 21) vs 2.85 (0.49) in the control group (*n* = 20); gaze 2.90 (0.30) in the intervention group (*n* = 21) vs 3.00 (0) in the control group (*n* = 20); silence 1.62 (0.80) in the intervention group (*n* = 21) vs 1.95 (0.83) in the control group (*n* = 20); expressions 2.43 (0.81) in the intervention group (*n* = 21) vs 2.45 (0.76) in the control group (*n* = 20); vocalisations 2.33 (0.73) in the intervention group (*n* = 21) vs 2.20 (0.70) in the control group (*n* = 20); infantised 2.38 (0.67) in the intervention group (*n* = 21) vs 2.20 (0.89) in the control group (*n* = 20); contingent response 2.57 (0.60) in the intervention group (*n* = 21) vs 2.25 (0.85) in the control group (*n* = 20); game playing 2.71 (0.46) in the intervention group (*n* = 21) vs 2.35 (0.75) in the control group (*n* = 20); summary 2.50 (0.29) in the intervention group (*n* = 21) vs 2.44 (0.43) in the control group (*n* = 20).

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			Mean mother–infant interactions (infant ratings): state 2.86 (0.35) in the intervention group (n = 22) vs 2.90 (0.31) in the control group (n = 20); activity 2.55 (0.60) in the intervention group (n = 22) vs 2.95 (0.22) in the control group (n = 20); orientation 2.00 (0.76) in the intervention group (n = 22) vs 2.85 (0.49) in the control group (n = 20); gaze 1.77 (0.81) in the intervention group (n = 22) vs 1.80 (0.95) in the control group (n = 20); expressions 2.45 (0.80) in the intervention group (n = 22) vs 2.50 (0.51) in the control group (n = 20); vocalisations 1.63 (0.73) in the intervention group (n = 22) vs 1.55 (0.69) in the control group (n = 20); fuzziness 2.32 (0.78) in the intervention group (n = 22) vs 2.50 (0.51) in the control group (n = 20); summary 2.23 (0.43) in the intervention group (n = 22) vs 2.39 (0.40) in the control group (n = 20). **Average number of sessions attended:** the intervention group received 2.8 CBT sessions (range 0–8), 0.6 child development sessions (range 0–3) and 1.8 social support building sessions (range 0–14). **Mean number of contacts overall:** 5.3 (range 0–24)
Milgrom *et al.*, 2005,[222] Australia RCT (between 5 and 10 women were recruited and then all randomised to the same treatment) 192 women	Community screening programme at 47 maternal and child health centres EPDS ≥ 12, DSM-IV diagnosis of depression, 37- to 42-week pregnancy, infant birthweight ≥ 2.5 kg, no congenital abnormality, no major health problem and no concurrent major psychiatric disorder	**Intervention:** Three intervention groups: group-based CBT, group-based counselling and individual counselling. The programmes consisted of nine weekly 90-minute sessions with mothers and three sessions involving partners. The group sessions were led by a senior therapist **Usual care:** Women were case managed by their maternal and child health nurse and referred to other agencies/services as necessary	**Any treatment vs RPC: mean difference in BDI scores:** complete case: 6.94, SE = 2.29, n = 121, p = 0.005, adjusted for baseline score complete case: 0.58, SE = 0.11, n = 121, p < 0.001; ITT: 6.94, SE = 2.29, n = 192, p = 0.005, adjusted for baseline score ITT: 0.58, SE = 0.11, n = 192, p < 0.001; **mean difference in BAI scores:** complete case: 7.86, SE = 2.25, n = 121, p < 0.01, adjusted for baseline score complete case: 0.54, SE = 0.08, n = 121, p < 0.001; ITT: 4.15, SE = 1.99, n = 192, p = 0.04, adjusted for baseline score ITT: 0.55, SE = 0.08, n = 192, p < 0.001. **CBT vs counselling: mean difference in BDI scores:** complete case: –0.065, SE = 1.86, n = 103, p = 0.97; ITT: –0.75, SE = 1.83, n = 159, p = 0.02; **mean difference in BAI scores:** complete case: 0.03, SE = 2.12, n = 103, p = 0.99; ITT: –0.25, SE = 1.80, n = 159, p = 0.89. **Group vs individual: mean difference in BDI scores:** complete case: 5.17, SE = 2.06, n = 103, p = 0.02; ITT: 2.94, SE = 1.99, n = 159, p = 0.14; **mean difference in BAI scores:** complete case: 2.41, SE = 2.46, n = 103, p = 0.34; ITT: 1.52, SE = 2.08, n = 159, p = 0.46.

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
			Threshold BDI at 3 months: 12 (35%) of the 34 women in the group counselling group, 16 (42%) of the 38 women in the individual counselling group, 14 (45%) of the 31 women in the CBT group and 13 (72%) of the 18 women in the control group had BDI scores of ≥ 17. **Mean BDI at 12 months:** CBT 12.17 (9.1) (n = 12), group counselling 13.77 (10.1) (n = 13), individual counselling 9.79 (5.9) (n = 24), control 21.13 (9.5) (n = 8). **Mean social provision score before and after 3 months intervention:** group CBT: baseline 70.88, SE = 3.00, follow-up 77.44, SE = 2.87 (n = 31); **group counselling:** baseline 68.96, SE = 2.21, follow-up 71.50, SE = 2.26 (n = 34); individual counselling: baseline 70.23, SE = 2.38, follow-up 76.37, SE = 1.98 (n = 38); control: baseline 73.00, SE = 3.22, follow-up 69.13, SE = 2.25 (n = 18)
Ingadottir and Thome, 2006,[210] Iceland Cluster controlled clinical trial 22 women	Six community health centres All postnatal mothers attending the clinics with an EPDS score of ≥ 12	**Intervention:** All nurses at the experimental centres attended a web-based distance learning programme on postnatal emotional distress and evidence-based interventions. The nurses at the experimental centres planned the nursing interventions individually for each woman and were supposed to use the evidence-based interventions that they had studied. They were also instructed to pay at least four home visits to each mother scoring ≥ 12 on the EPDS **Usual care:** Neither instructions on delivery of nursing care nor frequency of home visits were made to nurses at the control centres	**Threshold EPDS at 15 weeks:** 3 (25%) of the 12 women in the intervention group vs 9 (90%) of the 10 women in the control group had EPDS ≥ 12. **Threshold EPDS at 24 weeks:** 1 (8%) of the 12 women in the intervention group vs 3 (30%) of the 10 women in the control group had EPDS ≥ 12. **Mean EPDS at 15 weeks:** 9.4 in the intervention centres (n = 12) and 14.5 in the control centres (n = 10), p < 0.05. **Mean EPDS at 24 weeks:** 5.0 in the intervention centres (n = 8) and 10.4 in the control centres (n = 8), p < 0.05. **Mean Parent Stress Index (PSI-SF) at 15 weeks:** reduced from 75.2 to 62.4 for women in the intervention centres and from 80.7 to 73.3 for women at the control centres, p > 0.05; no significant differences were found in any of the subscales of the PSI-SF at the 15th week between the two groups with the exception of mother's distress (Mann–Whitney, p < 0.05). **Mean number of home visits:** 6.45 at experimental centres and 3.22 at control centres (z = –3.270, p < 0.05). **Nursing intervention:** active listening: 51 vs 16; emotional support: 43 vs 12; counselling: 39 vs 14; telephone consultation: 11 vs 4; support system enhancement: 7 vs 3; lactation counselling/teaching infant care: 4 vs 19; coping enhancement: 3 vs 7; health system guidance: 2 vs 9; teaching: disease process: 1 vs 4, socialisation enhancement: 2 vs 5

Characteristics	Setting, patient population	Intervention, usual care	Clinical outcomes (including timing of assessment)
Zlotnick *et al.*, 2006,[205] USA Individual RCT 99 women	Prenatal medical clinic Pregnant women at 23–32 weeks' gestation who were on public assistance, had a score of ≥ 27 on the Predictive Index for PND and who were not currently receiving mental health treatment	**Intervention:** Standard antenatal care + ROSE (Reach Out, Stand strong, Essentials for new mothers) programme intervention. The ROSE programme consists of 4 × 60-minute group sessions with three to five women over a 4-week period and a 50-minute individual booster session after delivery. **Enhancement:** therapists for the study were two nurses who received intensive training and supervision in delivering the intervention **Usual care:** Standard antenatal care alone	**Depression module of the longitudinal interval follow-up evaluation at 3 months:** 2 (4%) of the 46 women in the intervention group and 8 (20%) of the 40 women in the control group developed PND, $p = 0.04$. **Mean BDI at 3 months:** 9.39 (7.42) intervention ($n = 46$) vs 10.1 (9.41) control ($n = 40$). **Mean range of impaired functioning tool at 3 months:** 8.8 (2.58) intervention ($n = 46$) vs 10.2 (3.35) control ($n = 40$). Multivariate analysis of covariance found no significant difference between the two conditions in BDI score or the range of impaired functioning tool 3 months postpartum, controlling for baseline scores

ANOVA, analysis of variance; BAI, Beck Anxiety Inventory; BDI, Beck Depression Inventory; CBT, Cognitive Behavioural Therapy; CCEI, Crown–Crisp Experimental Index; DAS, Dyadic Adjustment Scale; DASS, Depression Anxiety Stress Scale; DSM, Diagnostic and Statistical Manual; EPDS, Edinburgh Postnatal Depression Scale; GHQ, General Health Questionnaire; HAMA, Hamilton Rating Scale for Anxiety; HAMD, Hamilton Rating Scale for Depression; HRSD, Hamilton Rating Sclae for Depression; ICD, International Classification of Diseases; ISEL, Interpersonal Support Evaluation List; ITT, intention to treat; LOCF, Last Observation Carried Forward; MADRS, Montgomery–Asberg Depression Rating Scale; MANOVA, multivariate analysis of variance; N/S, not significant; NSSQ TFS, Norbeck Social Support Questionnaire Total Functional Support; OCD, Obsessive Compulsive Disorder; OR, odds ratio; PANAS, Positive And Negative Affect Scale; PAS, Psychiatric Assessment Schedule; PMT, Pre-Menstrual Tension; POMS, Profile of Mood States; RCT, randomised controlled trial; RDC, Research Diagnostic Criteria; RPC, Routine Primary Care; SAS, Spielberger Trait/Anxiety Scale; SCID, Structured Clinical Interview for DSM; SCL-90-R, Symptom Checklist-90-Revised; SDA, Spanier Dyadic Adjustment Scale short form; SEI, Coopersmith's Self-Esteem Inventory; SSI, Social Support Interview; SSS, Sarason Social Support Scale; YBOCS, Yale-Brown Obsessive Compulsive Scale.

Appendix 6

Summary tables for update of NICE model of treatments for women with postnatal depression

Risk of discontinuation of treatment

Risk of discontinuation	Revised estimate			NICE estimate	
	Mean	SE	Distribution	Mean	Source
Absolute risk of discontinuation (usual care)	0.0861	0.0287	Normal	0.0664	NICE, 2007[30]
Relative risk of discontinuation (structured psychological therapy)	2.66	1.9783	Normal	2.66	NICE, 2007[30]
Relative risk of discontinuation (listening home visits)	1.49	0.8070	Normal	1.49	NICE, 2007[30]
NICE, National Institute for Health and Clinical Excellence; SE, standard error.					

Risk of no response/improvement

Risk of no response/improvement	Revised estimate			NICE estimate	
	Mean	SE	Distribution	Mean	Source
Absolute risk of no improvement (usual care)	0.6037	0.0514	Normal	0.6157	NICE, 2007[30]
Relative risk of no response (structured psychological therapy)	0.63	0.1276	Normal	0.63	NICE, 2007[30]
Relative risk of no response (listening home visits)	0.62	0.1119	Normal	0.62	NICE, 2007[30]
NICE, National Institute for Health and Clinical Excellence; SE, standard error.					

Risk of relapse

Risk of no response/improvement	Revised estimate			NICE estimate	
	Mean	SE	Distribution	Mean	Source
Absolute risk of relapse (usual care)	0.3120	0.0752	Normal	0.1957	NICE, 2004[89]
Relative risk of relapse (structured psychological therapy)	0.59	0.2525	Normal	1.00	NICE, 2004[89]
Relative risk of relapse (listening home visits)	0.59	0.2525	Normal	1.00	NICE, 2004[89]
NICE, National Institute for Health and Clinical Excellence; SE, standard error.					

Resource use

Treatment component	Resources used (per woman)	Source
Structured psychological therapy (treatment only)	Eight sessions×50 minutes with clinical psychologist	NICE, 2007[30]
Listening home visits (treatment only)	Eight home visits 45 minutes with health visitor	
Additional care	One community psychiatric nurse visit×1 hour, three GP visits×10 minutes, four health visitor home visits×45 minutes	

NICE, National Institute for Health and Clinical Excellence; SE, standard error.

Costs

	Revised estimate			NICE estimate	
Cost element	Value	Distribution	Source	Value	Source
Clinical psychologist (per hour)	£67.00	Fixed	Curtis, 2007[233]	£77	NICE, 2007[30]
Health visitor (per hour)	£91.00	Fixed		£89 (including travel)	
Health visitor (travel)	£1.30	Fixed			
GP (per hour)	£153.00	Fixed		£120	
Community psychiatric nurse (per hour)	£58.00	Fixed		£79 (including travel)	
Community psychiatric nurse (travel)	£1.30	Fixed			
Cost per treatment component	*Value*	*Distribution*	*Source*	*Value*	*Source*
Structured psychological therapy (treatment only)	£446.67	Fixed	NA	£513	NICE, 2007[30]
Listening home visits (treatment only)	£556.40	Fixed		£538	
Additional care	£414.00	Fixed		£408	
Total cost (per woman)	*Value*	*Distribution*	*Source*	*Value*	*Source*
Structured psychological therapy	£860.67	Fixed	NA	£921	NICE, 2007[30]
Listening home visits	£970.40	Fixed		£946	

NA, not applicable; NICE, National Institute for Health and Clinical Excellence; SE, standard error.

Utility weights

	Revised estimate			NICE estimate	
Utility weights	Mean	SE	Distribution	Mean	Source
Moderate depression	0.63	0.0275	Beta	0.63	Revicki and Wood, 1998[234]
Remission without maintenance treatment	0.86	0.0191	Beta	0.86	

NICE, National Institute for Health and Clinical Excellence; SE, standard error.

Health Technology Assessment reports published to date

Volume 1, 1997

No. 1
Home parenteral nutrition: a systematic review.
　By Richards DM, Deeks JJ, Sheldon TA, Shaffer JL.

No. 2
Diagnosis, management and screening of early localised prostate cancer.
　A review by Selley S, Donovan J, Faulkner A, Coast J, Gillatt D.

No. 3
The diagnosis, management, treatment and costs of prostate cancer in England and Wales.
　A review by Chamberlain J, Melia J, Moss S, Brown J.

No. 4
Screening for fragile X syndrome.
　A review by Murray J, Cuckle H, Taylor G, Hewison J.

No. 5
A review of near patient testing in primary care.
　By Hobbs FDR, Delaney BC, Fitzmaurice DA, Wilson S, Hyde CJ, Thorpe GH, *et al*.

No. 6
Systematic review of outpatient services for chronic pain control.
　By McQuay HJ, Moore RA, Eccleston C, Morley S, de C Williams AC.

No. 7
Neonatal screening for inborn errors of metabolism: cost, yield and outcome.
　A review by Pollitt RJ, Green A, McCabe CJ, Booth A, Cooper NJ, Leonard JV, *et al*.

No. 8
Preschool vision screening.
　A review by Snowdon SK, Stewart-Brown SL.

No. 9
Implications of socio-cultural contexts for the ethics of clinical trials.
　A review by Ashcroft RE, Chadwick DW, Clark SRL, Edwards RHT, Frith L, Hutton JL.

No. 10
A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment.
　By Davis A, Bamford J, Wilson I, Ramkalawan T, Forshaw M, Wright S.

No. 11
Newborn screening for inborn errors of metabolism: a systematic review.
　By Seymour CA, Thomason MJ, Chalmers RA, Addison GM, Bain MD, Cockburn F, *et al*.

No. 12
Routine preoperative testing: a systematic review of the evidence.
　By Munro J, Booth A, Nicholl J.

No. 13
Systematic review of the effectiveness of laxatives in the elderly.
　By Petticrew M, Watt I, Sheldon T.

No. 14
When and how to assess fast-changing technologies: a comparative study of medical applications of four generic technologies.
　A review by Mowatt G, Bower DJ, Brebner JA, Cairns JA, Grant AM, McKee L.

Volume 2, 1998

No. 1
Antenatal screening for Down's syndrome.
　A review by Wald NJ, Kennard A, Hackshaw A, McGuire A.

No. 2
Screening for ovarian cancer: a systematic review.
　By Bell R, Petticrew M, Luengo S, Sheldon TA.

No. 3
Consensus development methods, and their use in clinical guideline development.
　A review by Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, *et al*.

No. 4
A cost–utility analysis of interferon beta for multiple sclerosis.
　By Parkin D, McNamee P, Jacoby A, Miller P, Thomas S, Bates D.

No. 5
Effectiveness and efficiency of methods of dialysis therapy for end-stage renal disease: systematic reviews.
　By MacLeod A, Grant A, Donaldson C, Khan I, Campbell M, Daly C, *et al*.

No. 6
Effectiveness of hip prostheses in primary total hip replacement: a critical review of evidence and an economic model.
　By Faulkner A, Kennedy LG, Baxter K, Donovan J, Wilkinson M, Bevan G.

No. 7
Antimicrobial prophylaxis in colorectal surgery: a systematic review of randomised controlled trials.
　By Song F, Glenny AM.

No. 8
Bone marrow and peripheral blood stem cell transplantation for malignancy.
　A review by Johnson PWM, Simnett SJ, Sweetenham JW, Morgan GJ, Stewart LA.

No. 9
Screening for speech and language delay: a systematic review of the literature.
　By Law J, Boyle J, Harris F, Harkness A, Nye C.

No. 10
Resource allocation for chronic stable angina: a systematic review of effectiveness, costs and cost-effectiveness of alternative interventions.
　By Sculpher MJ, Petticrew M, Kelland JL, Elliott RA, Holdright DR, Buxton MJ.

No. 11
Detection, adherence and control of hypertension for the prevention of stroke: a systematic review.
　By Ebrahim S.

No. 12
Postoperative analgesia and vomiting, with special reference to day-case surgery: a systematic review.
　By McQuay HJ, Moore RA.

No. 13
Choosing between randomised and nonrandomised studies: a systematic review.
　By Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C.

No. 14
Evaluating patient-based outcome measures for use in clinical trials.
　A review by Fitzpatrick R, Davey C, Buxton MJ, Jones DR.

No. 15

Ethical issues in the design and conduct of randomised controlled trials.

A review by Edwards SJL, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J.

No. 16

Qualitative research methods in health technology assessment: a review of the literature.

By Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P.

No. 17

The costs and benefits of paramedic skills in pre-hospital trauma care.

By Nicholl J, Hughes S, Dixon S, Turner J, Yates D.

No. 18

Systematic review of endoscopic ultrasound in gastro-oesophageal cancer.

By Harris KM, Kelly S, Berry E, Hutton J, Roderick P, Cullingworth J, *et al*.

No. 19

Systematic reviews of trials and other studies.

By Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F.

No. 20

Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses.

A review by Fitzpatrick R, Shortall E, Sculpher M, Murray D, Morris R, Lodge M, *et al*.

Volume 3, 1999

No. 1

Informed decision making: an annotated bibliography and systematic review.

By Bekker H, Thornton JG, Airey CM, Connelly JB, Hewison J, Robinson MB, *et al*.

No. 2

Handling uncertainty when performing economic evaluation of healthcare interventions.

A review by Briggs AH, Gray AM.

No. 3

The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review.

By Crow R, Gage H, Hampson S, Hart J, Kimber A, Thomas H.

No. 4

A randomised controlled trial of different approaches to universal antenatal HIV testing: uptake and acceptability. Annex: Antenatal HIV testing – assessment of a routine voluntary approach.

By Simpson WM, Johnstone FD, Boyd FM, Goldberg DJ, Hart GJ, Gormley SM, *et al*.

No. 5

Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review.

By Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ.

No. 6

Assessing the costs of healthcare technologies in clinical trials.

A review by Johnston K, Buxton MJ, Jones DR, Fitzpatrick R.

No. 7

Cooperatives and their primary care emergency centres: organisation and impact.

By Hallam L, Henthorne K.

No. 8

Screening for cystic fibrosis.

A review by Murray J, Cuckle H, Taylor G, Littlewood J, Hewison J.

No. 9

A review of the use of health status measures in economic evaluation.

By Brazier J, Deverill M, Green C, Harper R, Booth A.

No. 10

Methods for the analysis of quality-of-life and survival data in health technology assessment.

A review by Billingham LJ, Abrams KR, Jones DR.

No. 11

Antenatal and neonatal haemoglobinopathy screening in the UK: review and economic analysis.

By Zeuner D, Ades AE, Karnon J, Brown J, Dezateux C, Anionwu EN.

No. 12

Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses.

A review by Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, *et al*.

No. 13

'Early warning systems' for identifying new healthcare technologies.

By Robert G, Stevens A, Gabbay J.

No. 14

A systematic review of the role of human papillomavirus testing within a cervical screening programme.

By Cuzick J, Sasieni P, Davies P, Adams J, Normand C, Frater A, *et al*.

No. 15

Near patient testing in diabetes clinics: appraising the costs and outcomes.

By Grieve R, Beech R, Vincent J, Mazurkiewicz J.

No. 16

Positron emission tomography: establishing priorities for health technology assessment.

A review by Robert G, Milne R.

No. 17 (Pt 1)

The debridement of chronic wounds: a systematic review.

By Bradley M, Cullum N, Sheldon T.

No. 17 (Pt 2)

Systematic reviews of wound care management: (2) Dressings and topical agents used in the healing of chronic wounds.

By Bradley M, Cullum N, Nelson EA, Petticrew M, Sheldon T, Torgerson D.

No. 18

A systematic literature review of spiral and electron beam computed tomography: with particular reference to clinical applications in hepatic lesions, pulmonary embolus and coronary artery disease.

By Berry E, Kelly S, Hutton J, Harris KM, Roderick P, Boyce JC, *et al*.

No. 19

What role for statins? A review and economic model.

By Ebrahim S, Davey Smith G, McCabe C, Payne N, Pickin M, Sheldon TA, *et al*.

No. 20

Factors that limit the quality, number and progress of randomised controlled trials.

A review by Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S, *et al*.

No. 21

Antimicrobial prophylaxis in total hip replacement: a systematic review.

By Glenny AM, Song F.

No. 22

Health promoting schools and health promotion in schools: two systematic reviews.

By Lister-Sharp D, Chapman S, Stewart-Brown S, Sowden A.

No. 23

Economic evaluation of a primary care-based education programme for patients with osteoarthritis of the knee.

A review by Lord J, Victor C, Littlejohns P, Ross FM, Axford JS.

Volume 4, 2000

No. 1
The estimation of marginal time preference in a UK-wide sample (TEMPUS) project.

A review by Cairns JA, van der Pol MM.

No. 2
Geriatric rehabilitation following fractures in older people: a systematic review.

By Cameron I, Crotty M, Currie C, Finnegan T, Gillespie L, Gillespie W, *et al*.

No. 3
Screening for sickle cell disease and thalassaemia: a systematic review with supplementary research.

By Davies SC, Cronin E, Gill M, Greengross P, Hickman M, Normand C.

No. 4
Community provision of hearing aids and related audiology services.

A review by Reeves DJ, Alborz A, Hickson FS, Bamford JM.

No. 5
False-negative results in screening programmes: systematic review of impact and implications.

By Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K.

No. 6
Costs and benefits of community postnatal support workers: a randomised controlled trial.

By Morrell CJ, Spiby H, Stewart P, Walters S, Morgan A.

No. 7
Implantable contraceptives (subdermal implants and hormonally impregnated intrauterine systems) versus other forms of reversible contraceptives: two systematic reviews to assess relative effectiveness, acceptability, tolerability and cost-effectiveness.

By French RS, Cowan FM, Mansour DJA, Morris S, Procter T, Hughes D, *et al*.

No. 8
An introduction to statistical methods for health technology assessment.

A review by White SJ, Ashby D, Brown PJ.

No. 9
Disease-modifying drugs for multiple sclerosis: a rapid and systematic review.

By Clegg A, Bryant J, Milne R.

No. 10
Publication and related biases.

A review by Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ.

No. 11
Cost and outcome implications of the organisation of vascular services.

By Michaels J, Brazier J, Palfreyman S, Shackley P, Slack R.

No. 12
Monitoring blood glucose control in diabetes mellitus: a systematic review.

By Coster S, Gulliford MC, Seed PT, Powrie JK, Swaminathan R.

No. 13
The effectiveness of domiciliary health visiting: a systematic review of international studies and a selective review of the British literature.

By Elkan R, Kendrick D, Hewitt M, Robinson JJA, Tolley K, Blair M, *et al*.

No. 14
The determinants of screening uptake and interventions for increasing uptake: a systematic review.

By Jepson R, Clegg A, Forbes C, Lewis R, Sowden A, Kleijnen J.

No. 15
The effectiveness and cost-effectiveness of prophylactic removal of wisdom teeth.

A rapid review by Song F, O'Meara S, Wilson P, Golder S, Kleijnen J.

No. 16
Ultrasound screening in pregnancy: a systematic review of the clinical effectiveness, cost-effectiveness and women's views.

By Bricker L, Garcia J, Henderson J, Mugford M, Neilson J, Roberts T, *et al*.

No. 17
A rapid and systematic review of the effectiveness and cost-effectiveness of the taxanes used in the treatment of advanced breast and ovarian cancer.

By Lister-Sharp D, McDonagh MS, Khan KS, Kleijnen J.

No. 18
Liquid-based cytology in cervical screening: a rapid and systematic review.

By Payne N, Chilcott J, McGoogan E.

No. 19
Randomised controlled trial of non-directive counselling, cognitive–behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care.

By King M, Sibbald B, Ward E, Bower P, Lloyd M, Gabbay M, *et al*.

No. 20
Routine referral for radiography of patients presenting with low back pain: is patients' outcome influenced by GPs' referral for plain radiography?

By Kerry S, Hilton S, Patel S, Dundas D, Rink E, Lord J.

No. 21
Systematic reviews of wound care management: (3) antimicrobial agents for chronic wounds; (4) diabetic foot ulceration.

By O'Meara S, Cullum N, Majid M, Sheldon T.

No. 22
Using routine data to complement and enhance the results of randomised controlled trials.

By Lewsey JD, Leyland AH, Murray GD, Boddy FA.

No. 23
Coronary artery stents in the treatment of ischaemic heart disease: a rapid and systematic review.

By Meads C, Cummins C, Jolly K, Stevens A, Burls A, Hyde C.

No. 24
Outcome measures for adult critical care: a systematic review.

By Hayes JA, Black NA, Jenkinson C, Young JD, Rowan KM, Daly K, *et al*.

No. 25
A systematic review to evaluate the effectiveness of interventions to promote the initiation of breastfeeding.

By Fairbank L, O'Meara S, Renfrew MJ, Woolridge M, Sowden AJ, Lister-Sharp D.

No. 26
Implantable cardioverter defibrillators: arrhythmias. A rapid and systematic review.

By Parkes J, Bryant J, Milne R.

No. 27
Treatments for fatigue in multiple sclerosis: a rapid and systematic review.

By Brañas P, Jordan R, Fry-Smith A, Burls A, Hyde C.

No. 28
Early asthma prophylaxis, natural history, skeletal development and economy (EASE): a pilot randomised controlled trial.

By Baxter-Jones ADG, Helms PJ, Russell G, Grant A, Ross S, Cairns JA, *et al*.

No. 29
Screening for hypercholesterolaemia versus case finding for familial hypercholesterolaemia: a systematic review and cost-effectiveness analysis.

By Marks D, Wonderling D, Thorogood M, Lambert H, Humphries SE, Neil HAW.

No. 30
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists in the medical management of unstable angina.

By McDonagh MS, Bachmann LM, Golder S, Kleijnen J, ter Riet G.

No. 31

A randomised controlled trial of prehospital intravenous fluid replacement therapy in serious trauma.

By Turner J, Nicholl J, Webber L, Cox H, Dixon S, Yates D.

No. 32

Intrathecal pumps for giving opioids in chronic pain: a systematic review.

By Williams JE, Louw G, Towlerton G.

No. 33

Combination therapy (interferon alfa and ribavirin) in the treatment of chronic hepatitis C: a rapid and systematic review.

By Shepherd J, Waugh N, Hewitson P.

No. 34

A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies.

By MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS.

No. 35

Intravascular ultrasound-guided interventions in coronary artery disease: a systematic literature review, with decision-analytic modelling, of outcomes and cost-effectiveness.

By Berry E, Kelly S, Hutton J, Lindsay HSJ, Blaxill JM, Evans JA, *et al*.

No. 36

A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression.

By Simpson S, Corney R, Fitzgerald P, Beecham J.

No. 37

Systematic review of treatments for atopic eczema.

By Hoare C, Li Wan Po A, Williams H.

No. 38

Bayesian methods in health technology assessment: a review.

By Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR.

No. 39

The management of dyspepsia: a systematic review.

By Delaney B, Moayyedi P, Deeks J, Innes M, Soo S, Barton P, *et al*.

No. 40

A systematic review of treatments for severe psoriasis.

By Griffiths CEM, Clark CM, Chalmers RJG, Li Wan Po A, Williams HC.

Volume 5, 2001

No. 1

Clinical and cost-effectiveness of donepezil, rivastigmine and galantamine for Alzheimer's disease: a rapid and systematic review.

By Clegg A, Bryant J, Nicholson T, McIntyre L, De Broe S, Gerard K, *et al*.

No. 2

The clinical effectiveness and cost-effectiveness of riluzole for motor neurone disease: a rapid and systematic review.

By Stewart A, Sandercock J, Bryan S, Hyde C, Barton PM, Fry-Smith A, *et al*.

No. 3

Equity and the economic evaluation of healthcare.

By Sassi F, Archard L, Le Grand J.

No. 4

Quality-of-life measures in chronic diseases of childhood.

By Eiser C, Morse R.

No. 5

Eliciting public preferences for healthcare: a systematic review of techniques.

By Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, *et al*.

No. 6

General health status measures for people with cognitive impairment: learning disability and acquired brain injury.

By Riemsma RP, Forbes CA, Glanville JM, Eastwood AJ, Kleijnen J.

No. 7

An assessment of screening strategies for fragile X syndrome in the UK.

By Pembrey ME, Barnicoat AJ, Carmichael B, Bobrow M, Turner G.

No. 8

Issues in methodological research: perspectives from researchers and commissioners.

By Lilford RJ, Richardson A, Stevens A, Fitzpatrick R, Edwards S, Rock F, *et al*.

No. 9

Systematic reviews of wound care management: (5) beds; (6) compression; (7) laser therapy, therapeutic ultrasound, electrotherapy and electromagnetic therapy.

By Cullum N, Nelson EA, Flemming K, Sheldon T.

No. 10

Effects of educational and psychosocial interventions for adolescents with diabetes mellitus: a systematic review.

By Hampson SE, Skinner TC, Hart J, Storey L, Gage H, Foxcroft D, *et al*.

No. 11

Effectiveness of autologous chondrocyte transplantation for hyaline cartilage defects in knees: a rapid and systematic review.

By Jobanputra P, Parry D, Fry-Smith A, Burls A.

No. 12

Statistical assessment of the learning curves of health technologies.

By Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT.

No. 13

The effectiveness and cost-effectiveness of temozolomide for the treatment of recurrent malignant glioma: a rapid and systematic review.

By Dinnes J, Cave C, Huang S, Major K, Milne R.

No. 14

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of debriding agents in treating surgical wounds healing by secondary intention.

By Lewis R, Whiting P, ter Riet G, O'Meara S, Glanville J.

No. 15

Home treatment for mental health problems: a systematic review.

By Burns T, Knapp M, Catty J, Healey A, Henderson J, Watt H, *et al*.

No. 16

How to develop cost-conscious guidelines.

By Eccles M, Mason J.

No. 17

The role of specialist nurses in multiple sclerosis: a rapid and systematic review.

By De Broe S, Christopher F, Waugh N.

No. 18

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of orlistat in the management of obesity.

By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

No. 19

The clinical effectiveness and cost-effectiveness of pioglitazone for type 2 diabetes mellitus: a rapid and systematic review.

By Chilcott J, Wight J, Lloyd Jones M, Tappenden P.

No. 20

Extended scope of nursing practice: a multicentre randomised controlled trial of appropriately trained nurses and preregistration house officers in preoperative assessment in elective general surgery.

By Kinley H, Czoski-Murray C, George S, McCabe C, Primrose J, Reilly C, *et al*.

No. 21
Systematic reviews of the effectiveness of day care for people with severe mental disorders: (1) Acute day hospital versus admission; (2) Vocational rehabilitation; (3) Day hospital versus outpatient care.
　By Marshall M, Crowther R, Almaraz- Serrano A, Creed F, Sledge W, Kluiter H, *et al*.

No. 22
The measurement and monitoring of surgical adverse events.
　By Bruce J, Russell EM, Mollison J, Krukowski ZH.

No. 23
Action research: a systematic review and guidance for assessment.
　By Waterman H, Tillen D, Dickson R, de Koning K.

No. 24
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of gemcitabine for the treatment of pancreatic cancer.
　By Ward S, Morris E, Bansback N, Calvert N, Crellin A, Forman D, *et al*.

No. 25
A rapid and systematic review of the evidence for the clinical effectiveness and cost-effectiveness of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer.
　By Lloyd Jones M, Hummel S, Bansback N, Orr B, Seymour M.

No. 26
Comparison of the effectiveness of inhaler devices in asthma and chronic obstructive airways disease: a systematic review of the literature.
　By Brocklebank D, Ram F, Wright J, Barry P, Cates C, Davies L, *et al*.

No. 27
The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint.
　By Bryan S, Weatherburn G, Bungay H, Hatrick C, Salas C, Parry D, *et al*.

No. 28
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of topotecan for ovarian cancer.
　By Forbes C, Shirran L, Bagnall A-M, Duffy S, ter Riet G.

No. 29
Superseded by a report published in a later volume.

No. 30
The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial.
　By Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M.

No. 31
Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients.
　By McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, *et al*.

No. 32
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of paclitaxel, docetaxel, gemcitabine and vinorelbine in non-small-cell lung cancer.
　By Clegg A, Scott DA, Sidhu M, Hewitson P, Waugh N.

No. 33
Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives.
　By Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G.

No. 34
Depot antipsychotic medication in the treatment of patients with schizophrenia: (1) Meta-review; (2) Patient and nurse attitudes.
　By David AS, Adams C.

No. 35
A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression.
　By Churchill R, Hunot V, Corney R, Knapp M, McGuire H, Tylee A, *et al*.

No. 36
Cost analysis of child health surveillance.
　By Sanderson D, Wright D, Acton C, Duree D.

Volume 6, 2002

No. 1
A study of the methods used to select review criteria for clinical audit.
　By Hearnshaw H, Harker R, Cheater F, Baker R, Grimshaw G.

No. 2
Fludarabine as second-line therapy for B cell chronic lymphocytic leukaemia: a technology assessment.
　By Hyde C, Wake B, Bryan S, Barton P, Fry-Smith A, Davenport C, *et al*.

No. 3
Rituximab as third-line treatment for refractory or recurrent Stage III or IV follicular non-Hodgkin's lymphoma: a systematic review and economic evaluation.
　By Wake B, Hyde C, Bryan S, Barton P, Song F, Fry-Smith A, *et al*.

No. 4
A systematic review of discharge arrangements for older people.
　By Parker SG, Peet SM, McPherson A, Cannaby AM, Baker R, Wilson A, *et al*.

No. 5
The clinical effectiveness and cost-effectiveness of inhaler devices used in the routine management of chronic asthma in older children: a systematic review and economic evaluation.
　By Peters J, Stevenson M, Beverley C, Lim J, Smith S.

No. 6
The clinical effectiveness and cost-effectiveness of sibutramine in the management of obesity: a technology assessment.
　By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

No. 7
The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: a systematic review.
　By Berry E, Kelly S, Westwood ME, Davies LM, Gough MJ, Bamford JM, *et al*.

No. 8
Promoting physical activity in South Asian Muslim women through 'exercise on prescription'.
　By Carroll B, Ali N, Azam N.

No. 9
Zanamivir for the treatment of influenza in adults: a systematic review and economic evaluation.
　By Burls A, Clark W, Stewart T, Preston C, Bryan S, Jefferson T, *et al*.

No. 10
A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models.
　By Richards RG, Sampson FC, Beard SM, Tappenden P.

No. 11
Screening for gestational diabetes: a systematic review and economic evaluation.
　By Scott DA, Loveman E, McIntyre L, Waugh N.

No. 12
The clinical effectiveness and cost-effectiveness of surgery for people with morbid obesity: a systematic review and economic evaluation.
　By Clegg AJ, Colquitt J, Sidhu MK, Royle P, Loveman E, Walker A.

No. 13
The clinical effectiveness of trastuzumab for breast cancer: a systematic review.
　By Lewis R, Bagnall A-M, Forbes C, Shirran E, Duffy S, Kleijnen J, *et al*.

No. 14
The clinical effectiveness and cost-effectiveness of vinorelbine for breast cancer: a systematic review and economic evaluation.
　By Lewis R, Bagnall A-M, King S, Woolacott N, Forbes C, Shirran L, *et al*.

213

No. 19
A rapid and systematic review and economic evaluation of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder.
By Bridle C, Palmer S, Bagnall A-M, Darba J, Duffy S, Sculpher M, *et al*.

No. 20
Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis.
By Karnon J, Peters J, Platt J, Chilcott J, McGoogan E, Brewer N.

No. 21
Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement.
By Avenell A, Broom J, Brown TJ, Poobalan A, Aucott L, Stearns SC, *et al*.

No. 22
Autoantibody testing in children with newly diagnosed type 1 diabetes mellitus.
By Dretzke J, Cummins C, Sandercock J, Fry-Smith A, Barrett T, Burls A.

No. 23
Clinical effectiveness and cost-effectiveness of prehospital intravenous fluids in trauma patients.
By Dretzke J, Sandercock J, Bayliss S, Burls A.

No. 24
Newer hypnotic drugs for the short-term management of insomnia: a systematic review and economic evaluation.
By Dündar Y, Boland A, Strobl J, Dodd S, Haycox A, Bagust A, *et al*.

No. 25
Development and validation of methods for assessing the quality of diagnostic accuracy studies.
By Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J.

No. 26
EVALUATE hysterectomy trial: a multicentre randomised trial comparing abdominal, vaginal and laparoscopic methods of hysterectomy.
By Garry R, Fountain J, Brown J, Manca A, Mason S, Sculpher M, *et al*.

No. 27
Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis.
By Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C.

No. 28
Effectiveness and cost-effectiveness of imatinib for first-line treatment of chronic myeloid leukaemia in chronic phase: a systematic review and economic analysis.
By Dalziel K, Round A, Stein K, Garside R, Price A.

No. 29
VenUS I: a randomised controlled trial of two types of bandage for treating venous leg ulcers.
By Iglesias C, Nelson EA, Cullum NA, Torgerson DJ, on behalf of the VenUS Team.

No. 30
Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction.
By Mowatt G, Vale L, Brazzelli M, Hernandez R, Murray A, Scott N, *et al*.

No. 31
A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme.
By Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S.

No. 32
The Social Support and Family Health Study: a randomised controlled trial and economic evaluation of two alternative forms of postnatal support for mothers living in disadvantaged inner-city areas.
By Wiggins M, Oakley A, Roberts I, Turner H, Rajan L, Austerberry H, *et al*.

No. 33
Psychosocial aspects of genetic screening of pregnant women and newborns: a systematic review.
By Green JM, Hewison J, Bekker HL, Bryant, Cuckle HS.

No. 34
Evaluation of abnormal uterine bleeding: comparison of three outpatient procedures within cohorts defined by age and menopausal status.
By Critchley HOD, Warner P, Lee AJ, Brechin S, Guise J, Graham B.

No. 35
Coronary artery stents: a rapid systematic review and economic evaluation.
By Hill R, Bagust A, Bakhai A, Dickson R, Dündar Y, Haycox A, *et al*.

No. 36
Review of guidelines for good practice in decision-analytic modelling in health technology assessment.
By Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al*.

No. 37
Rituximab (MabThera®) for aggressive non-Hodgkin's lymphoma: systematic review and economic evaluation.
By Knight C, Hind D, Brewer N, Abbott V.

No. 38
Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation.
By Jones L, Griffin S, Palmer S, Main C, Orton V, Sculpher M, *et al*.

No. 39
Pegylated interferon α-2a and -2b in combination with ribavirin in the treatment of chronic hepatitis C: a systematic review and economic evaluation.
By Shepherd J, Brodin H, Cave C, Waugh N, Price A, Gabbay J.

No. 40
Clopidogrel used in combination with aspirin compared with aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes: a systematic review and economic evaluation.
By Main C, Palmer S, Griffin S, Jones L, Orton V, Sculpher M, *et al*.

No. 41
Provision, uptake and cost of cardiac rehabilitation programmes: improving services to under-represented groups.
By Beswick AD, Rees K, Griebsch I, Taylor FC, Burke M, West RR, *et al*.

No. 42
Involving South Asian patients in clinical trials.
By Hussain-Gambles M, Leese B, Atkin K, Brown J, Mason S, Tovey P.

No. 43
Clinical and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes.
By Colquitt JL, Green C, Sidhu MK, Hartwell D, Waugh N.

No. 44
Identification and assessment of ongoing trials in health technology assessment reviews.
By Song FJ, Fry-Smith A, Davenport C, Bayliss S, Adi Y, Wilson JS, *et al*.

No. 45
Systematic review and economic evaluation of a long-acting insulin analogue, insulin glargine
By Warren E, Weatherley-Jones E, Chilcott J, Beverley C.

215

No. 46
Supplementation of a home-based exercise programme with a class-based programme for people with osteoarthritis of the knees: a randomised controlled trial and health economic analysis.

By McCarthy CJ, Mills PM, Pullen R, Richardson G, Hawkins N, Roberts CR, *et al*.

No. 47
Clinical and cost-effectiveness of once-daily versus more frequent use of same potency topical corticosteroids for atopic eczema: a systematic review and economic evaluation.

By Green C, Colquitt JL, Kirby J, Davidson P, Payne E.

No. 48
Acupuncture of chronic headache disorders in primary care: randomised controlled trial and economic analysis.

By Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, *et al*.

No. 49
Generalisability in economic evaluation studies in healthcare: a review and case studies.

By Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, *et al*.

No. 50
Virtual outreach: a randomised controlled trial and economic evaluation of joint teleconferenced medical consultations.

By Wallace P, Barber J, Clayton W, Currell R, Fleming K, Garner P, *et al*.

Volume 9, 2005

No. 1
Randomised controlled multiple treatment comparison to provide a cost-effectiveness rationale for the selection of antimicrobial therapy in acne.

By Ozolins M, Eady EA, Avery A, Cunliffe WJ, O'Neill C, Simpson NB, *et al*.

No. 2
Do the findings of case series studies vary significantly according to methodological characteristics?

By Dalziel K, Round A, Stein K, Garside R, Castelnuovo E, Payne L.

No. 3
Improving the referral process for familial breast cancer genetic counselling: findings of three randomised controlled trials of two interventions.

By Wilson BJ, Torrance N, Mollison J, Wordsworth S, Gray JR, Haites NE, *et al*.

No. 4
Randomised evaluation of alternative electrosurgical modalities to treat bladder outflow obstruction in men with benign prostatic hyperplasia.

By Fowler C, McAllister W, Plail R, Karim O, Yang Q.

No. 5
A pragmatic randomised controlled trial of the cost-effectiveness of palliative therapies for patients with inoperable oesophageal cancer.

By Shenfine J, McNamee P, Steen N, Bond J, Griffin SM.

No. 6
Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography.

By Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H.

No. 7
Issues in data monitoring and interim analysis of trials.

By Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, *et al*.

No. 8
Lay public's understanding of equipoise and randomisation in randomised controlled trials.

By Robinson EJ, Kerr CEP, Stevens AJ, Lilford RJ, Braunholtz DA, Edwards SJ, *et al*.

No. 9
Clinical and cost-effectiveness of electroconvulsive therapy for depressive illness, schizophrenia, catatonia and mania: systematic reviews and economic modelling studies.

By Greenhalgh J, Knight C, Hind D, Beverley C, Walters S.

No. 10
Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology.

By Smith SC, Lamping DL, Banerjee S, Harwood R, Foley B, Smith P, *et al*.

No. 11
Clinical effectiveness and cost-effectiveness of drotrecogin alfa (activated) (Xigris®) for the treatment of severe sepsis in adults: a systematic review and economic evaluation.

By Green C, Dinnes J, Takeda A, Shepherd J, Hartwell D, Cave C, *et al*.

No. 12
A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy.

By Dinnes J, Deeks J, Kirby J, Roderick P.

No. 13
Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK.

By Willis BH, Barton P, Pearmain P, Bryan S, Hyde C.

No. 14
Laparoscopic surgery for inguinal hernia repair: systematic review of effectiveness and economic evaluation.

By McCormack K, Wake B, Perez J, Fraser C, Cook J, McIntosh E, *et al*.

No. 15
Clinical effectiveness, tolerability and cost-effectiveness of newer drugs for epilepsy in adults: a systematic review and economic evaluation.

By Wilby J, Kainth A, Hawkins N, Epstein D, McIntosh H, McDaid C, *et al*.

No. 16
A randomised controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine.

By Peveler R, Kendrick T, Buxton M, Longworth L, Baldwin D, Moore M, *et al*.

No. 17
Clinical effectiveness and cost-effectiveness of immediate angioplasty for acute myocardial infarction: systematic review and economic evaluation.

By Hartwell D, Colquitt J, Loveman E, Clegg AJ, Brodin H, Waugh N, *et al*.

No. 18
A randomised controlled comparison of alternative strategies in stroke care.

By Kalra L, Evans A, Perez I, Knapp M, Swift C, Donaldson N.

No. 19
The investigation and analysis of critical incidents and adverse events in healthcare.

By Woloshynowych M, Rogers S, Taylor-Adams S, Vincent C.

No. 20
Potential use of routine databases in health technology assessment.

By Raftery J, Roderick P, Stevens A.

No. 21
Clinical and cost-effectiveness of newer immunosuppressive regimens in renal transplantation: a systematic review and modelling study.

By Woodroffe R, Yao GL, Meads C, Bayliss S, Ready A, Raftery J, *et al*.

No. 22
A systematic review and economic evaluation of alendronate, etidronate, risedronate, raloxifene and teriparatide for the prevention and treatment of postmenopausal osteoporosis.

By Stevenson M, Lloyd Jones M, De Nigris E, Brewer N, Davis S, Oakley J.

No. 23

A systematic review to examine the impact of psycho-educational interventions on health outcomes and costs in adults and children with difficult asthma.

By Smith JR, Mugford M, Holland R, Candy B, Noble MJ, Harrison BDW, *et al*.

No. 24

An evaluation of the costs, effectiveness and quality of renal replacement therapy provision in renal satellite units in England and Wales.

By Roderick P, Nicholson T, Armitage A, Mehta R, Mullee M, Gerard K, *et al*.

No. 25

Imatinib for the treatment of patients with unresectable and/or metastatic gastrointestinal stromal tumours: systematic review and economic evaluation.

By Wilson J, Connock M, Song F, Yao G, Fry-Smith A, Raftery J, *et al*.

No. 26

Indirect comparisons of competing interventions.

By Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, *et al*.

No. 27

Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling.

By Robinson M, Palmer S, Sculpher M, Philips Z, Ginnelly L, Bowens A, *et al*.

No. 28

Outcomes of electrically stimulated gracilis neosphincter surgery.

By Tillin T, Chambers M, Feldman R.

No. 29

The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation.

By Garside R, Stein K, Castelnuovo E, Pitt M, Ashcroft D, Dimmock P, *et al*.

No. 30

Systematic review on urine albumin testing for early detection of diabetic complications.

By Newman DJ, Mattock MB, Dawnay ABS, Kerry S, McGuire A, Yaqoob M, *et al*.

No. 31

Randomised controlled trial of the cost-effectiveness of water-based therapy for lower limb osteoarthritis.

By Cochrane T, Davey RC, Matthes Edwards SM.

No. 32

Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain.

By Thomas KJ, MacPherson H, Ratcliffe J, Thorpe L, Brazier J, Campbell M, *et al*.

No. 33

Cost-effectiveness and safety of epidural steroids in the management of sciatica.

By Price C, Arden N, Coglan L, Rogers P.

No. 34

The British Rheumatoid Outcome Study Group (BROSG) randomised controlled trial to compare the effectiveness and cost-effectiveness of aggressive versus symptomatic therapy in established rheumatoid arthritis.

By Symmons D, Tricker K, Roberts C, Davies L, Dawes P, Scott DL.

No. 35

Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials.

By King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, *et al*.

No. 36

The clinical and cost-effectiveness of implantable cardioverter defibrillators: a systematic review.

By Bryant J, Brodin H, Loveman E, Payne E, Clegg A.

No. 37

A trial of problem-solving by community mental health nurses for anxiety, depression and life difficulties among general practice patients. The CPN-GP study.

By Kendrick T, Simons L, Mynors-Wallis L, Gray A, Lathlean J, Pickering R, *et al*.

No. 38

The causes and effects of socio-demographic exclusions from clinical trials.

By Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, *et al*.

No. 39

Is hydrotherapy cost-effective? A randomised controlled trial of combined hydrotherapy programmes compared with physiotherapy land techniques in children with juvenile idiopathic arthritis.

By Epps H, Ginnelly L, Utley M, Southwood T, Gallivan S, Sculpher M, *et al*.

No. 40

A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study.

By Hobbs FDR, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S, *et al*.

No. 41

Displaced intracapsular hip fractures in fit, older people: a randomised comparison of reduction and fixation, bipolar hemiarthroplasty and total hip arthroplasty.

By Keating JF, Grant A, Masson M, Scott NW, Forbes JF.

No. 42

Long-term outcome of cognitive behaviour therapy clinical trials in central Scotland.

By Durham RC, Chambers JA, Power KG, Sharp DM, Macdonald RR, Major KA, *et al*.

No. 43

The effectiveness and cost-effectiveness of dual-chamber pacemakers compared with single-chamber pacemakers for bradycardia due to atrioventricular block or sick sinus syndrome: systematic review and economic evaluation.

By Castelnuovo E, Stein K, Pitt M, Garside R, Payne E.

No. 44

Newborn screening for congenital heart defects: a systematic review and cost-effectiveness analysis.

By Knowles R, Griebsch I, Dezateux C, Brown J, Bull C, Wren C.

No. 45

The clinical and cost-effectiveness of left ventricular assist devices for end-stage heart failure: a systematic review and economic evaluation.

By Clegg AJ, Scott DA, Loveman E, Colquitt J, Hutchinson J, Royle P, *et al*.

No. 46

The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma.

By Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D.

No. 47

Clinical and cost-effectiveness of autologous chondrocyte implantation for cartilage defects in knee joints: systematic review and economic evaluation.

By Clar C, Cummins E, McIntyre L, Thomas S, Lamb J, Bain L, *et al*.

217

No. 23
A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents.

By King S, Griffin S, Hodges Z, Weatherly H, Asseburg C, Richardson G, *et al.*

No. 24
The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review.

By Connock M, Burls A, Frew E, Fry-Smith A, Juarez-Garcia A, McCabe C, *et al.*

No. 25
Effectiveness and cost-effectiveness of salicylic acid and cryotherapy for cutaneous warts. An economic decision model.

By Thomas KS, Keogh-Brown MR, Chalmers JR, Fordham RJ, Holland RC, Armstrong SJ, *et al.*

No. 26
A systematic literature review of the effectiveness of non-pharmacological interventions to prevent wandering in dementia and evaluation of the ethical implications and acceptability of their use.

By Robinson L, Hutchings D, Corner L, Beyer F, Dickinson H, Vanoli A, *et al.*

No. 27
A review of the evidence on the effects and costs of implantable cardioverter defibrillator therapy in different patient groups, and modelling of cost-effectiveness and cost–utility for these groups in a UK context.

By Buxton M, Caine N, Chase D, Connelly D, Grace A, Jackson C, *et al.*

No. 28
Adefovir dipivoxil and pegylated interferon alfa-2a for the treatment of chronic hepatitis B: a systematic review and economic evaluation.

By Shepherd J, Jones J, Takeda A, Davidson P, Price A.

No. 29
An evaluation of the clinical and cost-effectiveness of pulmonary artery catheters in patient management in intensive care: a systematic review and a randomised controlled trial.

By Harvey S, Stevens K, Harrison D, Young D, Brampton W, McCabe C, *et al.*

No. 30
Accurate, practical and cost-effective assessment of carotid stenosis in the UK.

By Wardlaw JM, Chappell FM, Stevenson M, De Nigris E, Thomas S, Gillard J, *et al.*

No. 31
Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation.

By Woolacott N, Bravo Vergel Y, Hawkins N, Kainth A, Khadjesari Z, Misso K, *et al.*

No. 32
The cost-effectiveness of testing for hepatitis C in former injecting drug users.

By Castelnuovo E, Thompson-Coon J, Pitt M, Cramp M, Siebert U, Price A, *et al.*

No. 33
Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation.

By Kaltenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter M, Beverley C, *et al.*

No. 34
Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy.

By Williams C, Brunskill S, Altman D, Briggs A, Campbell H, Clarke M, *et al.*

No. 35
Psychological therapies including dialectical behaviour therapy for borderline personality disorder: a systematic review and preliminary economic evaluation.

By Brazier J, Tumur I, Holmes M, Ferriter M, Parry G, Dent-Brown K, *et al.*

No. 36
Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model.

By Whiting P, Westwood M, Bojke L, Palmer S, Richardson G, Cooper J, *et al.*

No. 37
Cognitive behavioural therapy in chronic fatigue syndrome: a randomised controlled trial of an outpatient group programme.

By O'Dowd H, Gladwell P, Rogers CA, Hollinghurst S, Gregory A.

No. 38
A comparison of the cost-effectiveness of five strategies for the prevention of nonsteroidal anti-inflammatory drug-induced gastrointestinal toxicity: a systematic review with economic modelling.

By Brown TJ, Hooper L, Elliott RA, Payne K, Webb R, Roberts C, et al.

No. 39
The effectiveness and cost-effectiveness of computed tomography screening for coronary artery disease: systematic review.

By Waugh N, Black C, Walker S, McIntyre L, Cummins E, Hillis G.

No. 40
What are the clinical outcome and cost-effectiveness of endoscopy undertaken by nurses when compared with doctors? A Multi-Institution Nurse Endoscopy Trial (MINuET).

By Williams J, Russell I, Durai D, Cheung W-Y, Farrin A, Bloor K, *et al.*

No. 41
The clinical and cost-effectiveness of oxaliplatin and capecitabine for the adjuvant treatment of colon cancer: systematic review and economic evaluation.

By Pandor A, Eggington S, Paisley S, Tappenden P, Sutcliffe P.

No. 42
A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness.

By Chen Y-F, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, *et al.*

No. 43
Telemedicine in dermatology: a randomised controlled trial.

By Bowns IR, Collins K, Walters SJ, McDonagh AJG.

No. 44
Cost-effectiveness of cell salvage and alternative methods of minimising perioperative allogeneic blood transfusion: a systematic review and economic model.

By Davies L, Brown TJ, Haynes S, Payne K, Elliott RA, McCollum C.

No. 45
Clinical effectiveness and cost-effectiveness of laparoscopic surgery for colorectal cancer: systematic reviews and economic evaluation.

By Murray A, Lourenco T, de Verteuil R, Hernandez R, Fraser C, McKinley A, *et al.*

No. 46
Etanercept and efalizumab for the treatment of psoriasis: a systematic review.

By Woolacott N, Hawkins N, Mason A, Kainth A, Khadjesari Z, Bravo Vergel Y, *et al.*

No. 47
Systematic reviews of clinical decision tools for acute abdominal pain.

By Liu JLY, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, *et al.*

No. 48
Evaluation of the ventricular assist device programme in the UK.

By Sharples L, Buxton M, Caine N, Cafferty F, Demiris N, Dyer M, *et al.*

219

No. 49

Cost-effectiveness of functional cardiac testing in the diagnosis and management of coronary artery disease: a randomised controlled trial. The CECaT trial.

By Sharples L, Hughes V, Crean A, Dyer M, Buxton M, Goldsmith K, *et al*.

No. 50

Evaluation of diagnostic tests when there is no gold standard. A review of methods.

By Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM.

No. 51

Systematic reviews of the clinical effectiveness and cost-effectiveness of proton pump inhibitors in acute upper gastrointestinal bleeding.

By Leontiadis GI, Sreedharan A, Dorward S, Barton P, Delaney B, Howden CW, *et al*.

No. 52

A review and critique of modelling in prioritising and designing screening programmes.

By Karnon J, Goyder E, Tappenden P, McPhie S, Towers I, Brazier J, *et al*.

No. 53

An assessment of the impact of the NHS Health Technology Assessment Programme.

By Hanney S, Buxton M, Green C, Coulson D, Raftery J.

Volume 12, 2008

No. 1

A systematic review and economic model of switching from nonglycopeptide to glycopeptide antibiotic prophylaxis for surgery.

By Cranny G, Elliott R, Weatherly H, Chambers D, Hawkins N, Myers L, *et al*.

No. 2

'Cut down to quit' with nicotine replacement therapies in smoking cessation: a systematic review of effectiveness and economic analysis.

By Wang D, Connock M, Barton P, Fry-Smith A, Aveyard P, Moore D.

No. 3

A systematic review of the effectiveness of strategies for reducing fracture risk in children with juvenile idiopathic arthritis with additional data on long-term risk of fracture and cost of disease management.

By Thornton J, Ashcroft D, O'Neill T, Elliott R, Adams J, Roberts C, *et al*.

No. 4

Does befriending by trained lay workers improve psychological well-being and quality of life for carers of people with dementia, and at what cost? A randomised controlled trial.

By Charlesworth G, Shepstone L, Wilson E, Thalanany M, Mugford M, Poland F.

No. 5

A multi-centre retrospective cohort study comparing the efficacy, safety and cost-effectiveness of hysterectomy and uterine artery embolisation for the treatment of symptomatic uterine fibroids. The HOPEFUL study.

By Hirst A, Dutton S, Wu O, Briggs A, Edwards C, Waldenmaier L, *et al*.

No. 6

Methods of prediction and prevention of pre-eclampsia: systematic reviews of accuracy and effectiveness literature with economic modelling.

By Meads CA, Cnossen JS, Meher S, Juarez-Garcia A, ter Riet G, Duley L, *et al*.

No. 7

The use of economic evaluations in NHS decision-making: a review and empirical investigation.

By Williams I, McIver S, Moore D, Bryan S.

No. 8

Stapled haemorrhoidectomy (haemorrhoidopexy) for the treatment of haemorrhoids: a systematic review and economic evaluation.

By Burch J, Epstein D, Baba-Akbari A, Weatherly H, Fox D, Golder S, *et al*.

No. 9

The clinical effectiveness of diabetes education models for Type 2 diabetes: a systematic review.

By Loveman E, Frampton GK, Clegg AJ.

No. 10

Payment to healthcare professionals for patient recruitment to trials: systematic review and qualitative study.

By Raftery J, Bryant J, Powell J, Kerr C, Hawker S.

No. 11

Cyclooxygenase-2 selective non-steroidal anti-inflammatory drugs (etodolac, meloxicam, celecoxib, rofecoxib, etoricoxib, valdecoxib and lumiracoxib) for osteoarthritis and rheumatoid arthritis: a systematic review and economic evaluation.

By Chen Y-F, Jobanputra P, Barton P, Bryan S, Fry-Smith A, Harris G, *et al*.

No. 12

The clinical effectiveness and cost-effectiveness of central venous catheters treated with anti-infective agents in preventing bloodstream infections: a systematic review and economic evaluation.

By Hockenhull JC, Dwan K, Boland A, Smith G, Bagust A, Dundar Y, *et al*.

No. 13

Stepped treatment of older adults on laxatives. The STOOL trial.

By Mihaylov S, Stark C, McColl E, Steen N, Vanoli A, Rubin G, *et al*.

No. 14

A randomised controlled trial of cognitive behaviour therapy in adolescents with major depression treated by selective serotonin reuptake inhibitors. The ADAPT trial.

By Goodyer IM, Dubicka B, Wilkinson P, Kelvin R, Roberts C, Byford S, *et al*.

No. 15

The use of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer: systematic review and economic evaluation.

By Hind D, Tappenden P, Tumur I, Eggington E, Sutcliffe P, Ryan A.

No. 16

Ranibizumab and pegaptanib for the treatment of age-related macular degeneration: a systematic review and economic evaluation.

By Colquitt JL, Jones J, Tan SC, Takeda A, Clegg AJ, Price A.

No. 17

Systematic review of the clinical effectiveness and cost-effectiveness of 64-slice or higher computed tomography angiography as an alternative to invasive coronary angiography in the investigation of coronary artery disease.

By Mowatt G, Cummins E, Waugh N, Walker S, Cook J, Jia X, *et al*.

No. 18

Structural neuroimaging in psychosis: a systematic review and economic evaluation.

By Albon E, Tsourapas A, Frew E, Davenport C, Oyebode F, Bayliss S, *et al*.

No. 19

Systematic review and economic analysis of the comparative effectiveness of different inhaled corticosteroids and their usage with long-acting beta$_2$ agonists for the treatment of chronic asthma in adults and children aged 12 years and over.

By Shepherd J, Rogers G, Anderson R, Main C, Thompson-Coon J, Hartwell D, *et al*.

No. 20

Systematic review and economic analysis of the comparative effectiveness of different inhaled corticosteroids and their usage with long-acting beta$_2$ agonists for the treatment of chronic asthma in children under the age of 12 years.

By Main C, Shepherd J, Anderson R, Rogers G, Thompson-Coon J, Liu Z, *et al*.

No. 21

Ezetimibe for the treatment of hypercholesterolaemia: a systematic review and economic evaluation.

By Ara R, Tumur I, Pandor A, Duenas A, Williams R, Wilkinson A, *et al*.

No. 22

Topical or oral ibuprofen for chronic knee pain in older people. The TOIB study.

By Underwood M, Ashby D, Carnes D, Castelnuovo E, Cross P, Harding G, *et al*.

No. 23

A prospective randomised comparison of minor surgery in primary and secondary care. The MiSTIC trial.

By George S, Pockney P, Primrose J, Smith H, Little P, Kinley H, *et al*.

No. 24

A review and critical appraisal of measures of therapist–patient interactions in mental health settings.

By Cahill J, Barkham M, Hardy G, Gilbody S, Richards D, Bower P, *et al*.

No. 25

The clinical effectiveness and cost-effectiveness of screening programmes for amblyopia and strabismus in children up to the age of 4–5 years: a systematic review and economic evaluation.

By Carlton J, Karnon J, Czoski-Murray C, Smith KJ, Marr J.

No. 26

A systematic review of the clinical effectiveness and cost-effectiveness and economic modelling of minimal incision total hip replacement approaches in the management of arthritic disease of the hip.

By de Verteuil R, Imamura M, Zhu S, Glazener C, Fraser C, Munro N, *et al*.

No. 27

A preliminary model-based assessment of the cost–utility of a screening programme for early age-related macular degeneration.

By Karnon J, Czoski-Murray C, Smith K, Brand C, Chakravarthy U, Davis S, *et al*.

No. 28

Intravenous magnesium sulphate and sotalol for prevention of atrial fibrillation after coronary artery bypass surgery: a systematic review and economic evaluation.

By Shepherd J, Jones J, Frampton GK, Tanajewski L, Turner D, Price A.

No. 29

Absorbent products for urinary/faecal incontinence: a comparative evaluation of key product categories.

By Fader M, Cottenden A, Getliffe K, Gage H, Clarke-O'Neill S, Jamieson K, *et al*.

No. 30

A systematic review of repetitive functional task practice with modelling of resource use, costs and effectiveness.

By French B, Leathley M, Sutton C, McAdam J, Thomas L, Forster A, *et al*.

No. 31

The effectiveness and cost-effectivness of minimal access surgery amongst people with gastro-oesophageal reflux disease – a UK collaborative study. The REFLUX trial.

By Grant A, Wileman S, Ramsay C, Bojke L, Epstein D, Sculpher M, *et al*.

No. 32

Time to full publication of studies of anti-cancer medicines for breast cancer and the potential for publication bias: a short systematic review.

By Takeda A, Loveman E, Harris P, Hartwell D, Welch K.

No. 33

Performance of screening tests for child physical abuse in accident and emergency departments.

By Woodman J, Pitt M, Wentz R, Taylor B, Hodes D, Gilbert RE.

No. 34

Curative catheter ablation in atrial fibrillation and typical atrial flutter: systematic review and economic evaluation.

By Rodgers M, McKenna C, Palmer S, Chambers D, Van Hout S, Golder S, *et al*.

No. 35

Systematic review and economic modelling of effectiveness and cost utility of surgical treatments for men with benign prostatic enlargement.

By Lourenco T, Armstrong N, N'Dow J, Nabi G, Deverill M, Pickard R, *et al*.

No. 36

Immunoprophylaxis against respiratory syncytial virus (RSV) with palivizumab in children: a systematic review and economic evaluation.

By Wang D, Cummins C, Bayliss S, Sandercock J, Burls A.

Volume 13, 2009

No. 1

Deferasirox for the treatment of iron overload associated with regular blood transfusions (transfusional haemosiderosis) in patients suffering with chronic anaemia: a systematic review and economic evaluation.

By McLeod C, Fleeman N, Kirkham J, Bagust A, Boland A, Chu P, *et al*.

No. 2

Thrombophilia testing in people with venous thromboembolism: systematic review and cost-effectiveness analysis.

By Simpson EL, Stevenson MD, Rawdin A, Papaioannou D.

No. 3

Surgical procedures and non-surgical devices for the management of non-apnoeic snoring: a systematic review of clinical effects and associated treatment costs.

By Main C, Liu Z, Welch K, Weiner G, Quentin Jones S, Stein K.

No. 4

Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea–hypopnoea syndrome: a systematic review and economic analysis.

By McDaid C, Griffin S, Weatherly H, Durée K, van der Burgt M, van Hout S, Akers J, *et al*.

No. 5

Use of classical and novel biomarkers as prognostic risk factors for localised prostate cancer: a systematic review.

By Sutcliffe P, Hummel S, Simpson E, Young T, Rees A, Wilkinson A, *et al*.

No. 6

The harmful health effects of recreational ecstasy: a systematic review of observational evidence.

By Rogers G, Elston J, Garside R, Roome C, Taylor R, Younger P, *et al*.

No. 7

Systematic review of the clinical effectiveness and cost-effectiveness of oesophageal Doppler monitoring in critically ill and high-risk surgical patients.

By Mowatt G, Houston G, Hernández R, de Verteuil R, Fraser C, Cuthbertson B, *et al*.

No. 8

The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports.

By Taylor RS, Elston J.

No. 9

Controlling Hypertension and Hypotension Immediately Post Stroke (CHHIPS) – a randomised controlled trial.

By Potter J, Mistri A, Brodie F, Chernova J, Wilson E, Jagger C, *et al*.

No. 30
Psychological interventions for postnatal depression: cluster randomised trial and economic evaluation. The PoNDER trial.
By Morrell CJ, Warner R, Slade P, Dixon S, Walters S, Paley G, *et al*.

No. 31
The effect of different treatment durations of clopidogrel in patients with non-ST-segment elevation acute coronary syndromes: a systematic review and value of information analysis.
By Rogowski R, Burch J, Palmer S, Craigs C, Golder S, Woolacott N.

No. 32
Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care.
By Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, *et al*.

No. 33
A multicentre randomised controlled trial of the use of continuous positive airway pressure and non-invasive positive pressure ventilation in the early treatment of patients presenting to the emergency department with severe acute cardiogenic pulmonary oedema: the 3CPO trial.
By Gray AJ, Goodacre S, Newby DE, Masson MA, Sampson F, Dixon S, *et al*., on behalf of the 3CPO study investigators.

No. 34
Early high-dose lipid-lowering therapy to avoid cardiac events: a systematic review and economic evaluation.
By Ara R, Pandor A, Stevens J, Rees A, Rafia R.

No. 35
Adefovir dipivoxil and pegylated interferon alpha for the treatment of chronic hepatitis B: an updated systematic review and economic evaluation.
By Jones J, Shepherd J, Baxter L, Gospodarevskaya E, Hartwell D, Harris P, *et al*.

225

Health Technology Assessment
programme

Director,
Professor Tom Walley,
Director, NIHR HTA
programme, Professor of
Clinical Pharmacology,
University of Liverpool

Deputy Director,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley,
Director, NIHR HTA
programme, Professor of
Clinical Pharmacology,
University of Liverpool

Deputy Chair,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield

Dr Bob Coates,
Consultant Advisor, NETSCC,
HTA

Dr Andrew Cook,
Consultant Advisor, NETSCC,
HTA

Dr Peter Davidson,
Director of Science Support,
NETSCC, HTA

Professor Robin E Ferner,
Consultant Physician and
Director, West Midlands Centre
for Adverse Drug Reactions,
City Hospital NHS Trust,
Birmingham

Professor Paul Glasziou,
Professor of Evidence-Based
Medicine, University of Oxford

Dr Nick Hicks,
Director of NHS Support,
NETSCC, HTA

Dr Edmund Jessop,
Medical Adviser, National
Specialist, National
Commissioning Group (NCG),
Department of Health, London

Ms Lynn Kerridge,
Chief Executive Officer,
NETSCC and NETSCC, HTA

Dr Ruairidh Milne,
Director of Strategy and
Development, NETSCC

Ms Kay Pattison,
Section Head, NHS R&D
Programme, Department of
Health

Ms Pamela Young,
Specialist Programme Manager,
NETSCC, HTA

HTA Commissioning Board

Members

Programme Director,
Professor Tom Walley,
Director, NIHR HTA
programme, Professor of
Clinical Pharmacology,
University of Liverpool

Chair,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield

Deputy Chair,
Dr Andrew Farmer,
Senior Lecturer in General
Practice, Department of
Primary Health Care,
University of Oxford

Professor Ann Ashburn,
Professor of Rehabilitation
and Head of Research,
Southampton General Hospital

Professor Deborah Ashby,
Professor of Medical Statistics,
Queen Mary, University of
London

Professor John Cairns,
Professor of Health Economics,
London School of Hygiene and
Tropical Medicine

Professor Peter Croft,
Director of Primary Care
Sciences Research Centre, Keele
University

Professor Nicky Cullum,
Director of Centre for Evidence-
Based Nursing, University of
York

Professor Jenny Donovan,
Professor of Social Medicine,
University of Bristol

Professor Steve Halligan,
Professor of Gastrointestinal
Radiology, University College
Hospital, London

Professor Freddie Hamdy,
Professor of Urology,
University of Sheffield

Professor Allan House,
Professor of Liaison Psychiatry,
University of Leeds

Dr Martin J Landray,
Reader in Epidemiology,
Honorary Consultant Physician,
Clinical Trial Service Unit,
University of Oxford

Professor Stuart Logan,
Director of Health & Social
Care Research, The Peninsula
Medical School, Universities of
Exeter and Plymouth

Dr Rafael Perera,
Lecturer in Medical Statisitics,
Department of Primary Health
Care, Univeristy of Oxford

Professor Ian Roberts,
Professor of Epidemiology &
Public Health, London School
of Hygiene and Tropical
Medicine

Professor Mark Sculpher,
Professor of Health Economics,
University of York

Professor Helen Smith,
Professor of Primary Care,
University of Brighton

Professor Kate Thomas,
Professor of Complementary &
Alternative Medicine Research,
University of Leeds

Professor David John
Torgerson,
Director of York Trials Unit,
University of York

Professor Hywel Williams,
Professor of Dermato-
Epidemiology, University of
Nottingham

Observers

Ms Kay Pattison,
Section Head, NHS R&D
Programme, Department of
Health

Dr Morven Roberts,
Clinical Trials Manager,
Medical Research Council

Diagnostic Technologies & Screening Panel

Members

Chair,
Professor Paul Glasziou,
Professor of Evidence-Based
Medicine, University of Oxford

Deputy Chair,
Dr David Elliman,
Consultant Paediatrician and
Honorary Senior Lecturer,
Great Ormond Street Hospital,
London

Professor Judith E Adams,
Consultant Radiologist,
Manchester Royal Infirmary,
Central Manchester &
Manchester Children's
University Hospitals NHS Trust,
and Professor of Diagnostic
Radiology, Imaging Science
and Biomedical Engineering,
Cancer & Imaging Sciences,
University of Manchester

Ms Jane Bates,
Consultant Ultrasound
Practitioner, Ultrasound
Department, Leeds Teaching
Hospital NHS Trust

Dr Stephanie Dancer,
Consultant Microbiologist,
Hairmyres Hospital, East
Kilbride

Professor Glyn Elwyn,
Primary Medical Care Research
Group, Swansea Clinical School,
University of Wales

Dr Ron Gray,
Consultant Clinical
Epidemiologist, Department
of Public Health, University of
Oxford

Professor Paul D Griffiths,
Professor of Radiology,
University of Sheffield

Dr Jennifer J Kurinczuk,
Consultant Clinical
Epidemiologist, National
Perinatal Epidemiology Unit,
Oxford

Dr Susanne M Ludgate,
Medical Director, Medicines &
Healthcare Products Regulatory
Agency, London

Dr Anne Mackie,
Director of Programmes, UK
National Screening Committee

Dr Michael Millar,
Consultant Senior Lecturer in
Microbiology, Barts and The
London NHS Trust, Royal
London Hospital

Mr Stephen Pilling,
Director, Centre for Outcomes,
Research & Effectiveness,
Joint Director, National
Collaborating Centre for
Mental Health, University
College London

Mrs Una Rennard,
Service User Representative

Dr Phil Shackley,
Senior Lecturer in Health
Economics, School of
Population and Health
Sciences, University of
Newcastle upon Tyne

Dr W Stuart A Smellie,
Consultant in Chemical
Pathology, Bishop Auckland
General Hospital

Dr Nicholas Summerton,
Consultant Clinical and Public
Health Advisor, NICE

Ms Dawn Talbot,
Service User Representative

Dr Graham Taylor,
Scientific Advisor, Regional
DNA Laboratory, St James's
University Hospital, Leeds

Professor Lindsay Wilson
Turnbull,
Scientific Director of the
Centre for Magnetic Resonance
Investigations and YCR
Professor of Radiology, Hull
Royal Infirmary

Observers

Dr Tim Elliott,
Team Leader, Cancer
Screening, Department of
Health

Dr Catherine Moody,
Programme Manager,
Neuroscience and Mental
Health Board

Dr Ursula Wells,
Principal Research Officer,
Department of Health

Pharmaceuticals Panel

Members

Chair,
Professor Robin Ferner,
Consultant Physician and
Director, West Midlands Centre
for Adverse Drug Reactions,
City Hospital NHS Trust,
Birmingham

Deputy Chair,
Professor Imti Choonara,
Professor in Child Health,
University of Nottingham

Mrs Nicola Carey,
Senior Research Fellow,
School of Health and Social
Care, The University of
Reading

Mr John Chapman,
Service User Representative

Dr Peter Elton,
Director of Public Health,
Bury Primary Care Trust

Dr Ben Goldacre,
Research Fellow, Division of
Psychological Medicine and
Psychiatry, King's College
London

Mrs Barbara Greggains,
Service User Representative

Dr Bill Gutteridge,
Medical Adviser, London
Strategic Health Authority

Dr Dyfrig Hughes,
Reader in Pharmacoeconomics
and Deputy Director, Centre
for Economics and Policy in
Health, IMSCaR, Bangor
University

Professor Jonathan Ledermann,
Professor of Medical Oncology
and Director of the Cancer
Research UK and University
College London Cancer Trials
Centre

Dr Yoon K Loke,
Senior Lecturer in Clinical
Pharmacology, University of
East Anglia

Professor Femi Oyebode,
Consultant Psychiatrist
and Head of Department,
University of Birmingham

Dr Andrew Prentice,
Senior Lecturer and Consultant
Obstetrician and Gynaecologist,
The Rosie Hospital, University
of Cambridge

Dr Martin Shelly,
General Practitioner, Leeds,
and Associate Director, NHS
Clinical Governance Support
Team, Leicester

Dr Gillian Shepherd,
Director, Health and Clinical
Excellence, Merck Serono Ltd

Mrs Katrina Simister,
Assistant Director New
Medicines, National Prescribing
Centre, Liverpool

Mr David Symes,
Service User Representative

Dr Lesley Wise,
Unit Manager,
Pharmacoepidemiology
Research Unit, VRMM,
Medicines & Healthcare
Products Regulatory Agency

Observers

Ms Kay Pattison,
Section Head, NHS R&D
Programme, Department of
Health

Mr Simon Reeve,
Head of Clinical and Cost-
Effectiveness, Medicines,
Pharmacy and Industry Group,
Department of Health

Dr Heike Weber,
Programme Manager,
Medical Research Council

Dr Ursula Wells,
Principal Research Officer,
Department of Health

Current and past membership details of all HTA programme 'committees' are available from the HTA website (www.hta.ac.uk)

Therapeutic Procedures Panel

Disease Prevention Panel

Expert Advisory Network

Members

Professor Douglas Altman,
Professor of Statistics in
Medicine, Centre for Statistics
in Medicine, University of
Oxford

Professor John Bond,
Professor of Social Gerontology
& Health Services Research,
University of Newcastle upon
Tyne

Professor Andrew Bradbury,
Professor of Vascular Surgery,
Solihull Hospital, Birmingham

Mr Shaun Brogan,
Chief Executive, Ridgeway
Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,
Chief Executive, Regulation
and Improvement Authority,
Belfast

Ms Tracy Bury,
Project Manager, World
Confederation for Physical
Therapy, London

Professor Iain T Cameron,
Professor of Obstetrics and
Gynaecology and Head of the
School of Medicine, University
of Southampton

Dr Christine Clark,
Medical Writer and Consultant
Pharmacist, Rossendale

Professor Collette Clifford,
Professor of Nursing and
Head of Research, The
Medical School, University of
Birmingham

Professor Barry Cookson,
Director, Laboratory of Hospital
Infection, Public Health
Laboratory Service, London

Dr Carl Counsell,
Clinical Senior Lecturer in
Neurology, University of
Aberdeen

Professor Howard Cuckle,
Professor of Reproductive
Epidemiology, Department
of Paediatrics, Obstetrics &
Gynaecology, University of
Leeds

Dr Katherine Darton,
Information Unit, MIND – The
Mental Health Charity, London

Professor Carol Dezateux,
Professor of Paediatric
Epidemiology, Institute of Child
Health, London

Mr John Dunning,
Consultant Cardiothoracic
Surgeon, Papworth Hospital
NHS Trust, Cambridge

Mr Jonothan Earnshaw,
Consultant Vascular Surgeon,
Gloucestershire Royal Hospital,
Gloucester

Professor Martin Eccles,
Professor of Clinical
Effectiveness, Centre for Health
Services Research, University of
Newcastle upon Tyne

Professor Pam Enderby,
Dean of Faculty of Medicine,
Institute of General Practice
and Primary Care, University of
Sheffield

Professor Gene Feder,
Professor of Primary Care
Research & Development,
Centre for Health Sciences,
Barts and The London School
of Medicine and Dentistry

Mr Leonard R Fenwick,
Chief Executive, Freeman
Hospital, Newcastle upon Tyne

Mrs Gillian Fletcher,
Antenatal Teacher and Tutor
and President, National
Childbirth Trust, Henfield

Professor Jayne Franklyn,
Professor of Medicine,
University of Birmingham

Mr Tam Fry,
Honorary Chairman, Child
Growth Foundation, London

Professor Fiona Gilbert,
Consultant Radiologist and
NCRN Member, University of
Aberdeen

Professor Paul Gregg,
Professor of Orthopaedic
Surgical Science, South Tees
Hospital NHS Trust

Bec Hanley,
Co-director, TwoCan Associates,
West Sussex

Dr Maryann L Hardy,
Senior Lecturer, University of
Bradford

Mrs Sharon Hart,
Healthcare Management
Consultant, Reading

Professor Robert E Hawkins,
CRC Professor and Director
of Medical Oncology, Christie
CRC Research Centre,
Christie Hospital NHS Trust,
Manchester

Professor Richard Hobbs,
Head of Department of Primary
Care & General Practice,
University of Birmingham

Professor Alan Horwich,
Dean and Section Chairman,
The Institute of Cancer
Research, London

Professor Allen Hutchinson,
Director of Public Health and
Deputy Dean of ScHARR,
University of Sheffield

Professor Peter Jones,
Professor of Psychiatry,
University of Cambridge,
Cambridge

Professor Stan Kaye,
Cancer Research UK Professor
of Medical Oncology, Royal
Marsden Hospital and Institute
of Cancer Research, Surrey

Dr Duncan Keeley,
General Practitioner (Dr Burch
& Ptnrs), The Health Centre,
Thame

Dr Donna Lamping,
Research Degrees Programme
Director and Reader in
Psychology, Health Services
Research Unit, London School
of Hygiene and Tropical
Medicine, London

Mr George Levvy,
Chief Executive, Motor
Neurone Disease Association,
Northampton

Professor James Lindesay,
Professor of Psychiatry for the
Elderly, University of Leicester

Professor Julian Little,
Professor of Human Genome
Epidemiology, University of
Ottawa

Professor Alistaire McGuire,
Professor of Health Economics,
London School of Economics

Professor Rajan Madhok,
Medical Director and Director
of Public Health, Directorate
of Clinical Strategy & Public
Health, North & East Yorkshire
& Northern Lincolnshire
Health Authority, York

Professor Alexander Markham,
Director, Molecular Medicine
Unit, St James's University
Hospital, Leeds

Dr Peter Moore,
Freelance Science Writer,
Ashtead

Dr Andrew Mortimore,
Public Health Director,
Southampton City Primary
Care Trust

Dr Sue Moss,
Associate Director, Cancer
Screening Evaluation Unit,
Institute of Cancer Research,
Sutton

Professor Miranda Mugford,
Professor of Health Economics
and Group Co-ordinator,
University of East Anglia

Professor Jim Neilson,
Head of School of Reproductive
& Developmental Medicine
and Professor of Obstetrics
and Gynaecology, University of
Liverpool

Mrs Julietta Patnick,
National Co-ordinator, NHS
Cancer Screening Programmes,
Sheffield

Professor Robert Peveler,
Professor of Liaison Psychiatry,
Royal South Hants Hospital,
Southampton

Professor Chris Price,
Director of Clinical Research,
Bayer Diagnostics Europe,
Stoke Poges

Professor William Rosenberg,
Professor of Hepatology
and Consultant Physician,
University of Southampton

Professor Peter Sandercock,
Professor of Medical Neurology,
Department of Clinical
Neurosciences, University of
Edinburgh

Dr Susan Schonfield,
Consultant in Public Health,
Hillingdon Primary Care Trust,
Middlesex

Dr Eamonn Sheridan,
Consultant in Clinical Genetics,
St James's University Hospital,
Leeds

Dr Margaret Somerville,
Director of Public Health
Learning, Peninsula Medical
School, University of Plymouth

Professor Sarah Stewart-Brown,
Professor of Public Health,
Division of Health in the
Community, University of
Warwick, Coventry

Professor Ala Szczepura,
Professor of Health Service
Research, Centre for Health
Services Studies, University of
Warwick, Coventry

Mrs Joan Webster,
Consumer Member, Southern
Derbyshire Community Health
Council

Professor Martin Whittle,
Clinical Co-director, National
Co-ordinating Centre for
Women's and Children's
Health, Lymington

230