BMJ

Screening for postnatal depression in primary care: cost effectiveness analysis

Mike Paulden, research fellow,[1] Stephen Palmer, senior research fellow,[1] Catherine Hewitt, research fellow,[2] Simon Gilbody, professor of psychological medicine and health services research[2]

[1]Centre for Health Economics, University of York, Heslington, York YO10 5DD

[2]Department of Health Sciences, University of York

Correspondence to: M Paulden, Toronto Health Economics and Technology Assessment (THETA) Collaborative, University of Toronto, Leslie Dan Pharmacy Building, 144 College Street, Toronto, Ontario, Canada, M5S 3M2 mike.paulden@theta.utoronto.ca

ABSTRACT

Objective To evaluate the cost effectiveness of routine screening for postnatal depression in primary care.

Design Cost effectiveness analysis with a decision model of alternative methods of screening for depression, including standardised postnatal depression and generic depression instruments. The performance of screening instruments was derived from a systematic review and bivariate meta-analysis at a range of instrument cut points; estimates of other relevant parameters were derived from literature sources and relevant databases. A decision tree considered the full treatment pathway from the possible onset of postnatal depression through identification, treatment, and possible relapse.

Setting Primary care.

Participants A hypothetical population of women assessed for postnatal depression either via routine care only or supplemented by use of formal identification methods six weeks postnatally, as recommended in recent guidelines.

Main outcome measures Costs expressed in 2006-7 prices and impact on health outcomes expressed in terms of quality adjusted life years (QALYs). The time horizon of the analysis was one year.

Results The routine application of either postnatal or general depression questionnaires did not seem to be cost effective compared with routine care only. The Edinburgh postnatal depression scale (at a cut point of 16) had an incremental cost effectiveness ratio (ICER) of £41 103 (€45 398, $67 130) per QALY compared with routine care only. The ICER for all other strategies ranged from £49 928 to £272 463 per QALY versus routine care only, while the probability that no formal identification strategy was cost effective was 88% (59%) at a cost effectiveness threshold of £20 000 (£30 000) per QALY. While sensitivity analysis indicated that the cost of managing incorrectly identified depression (false positive result) was an important driver of the model, formal identification approaches did not seem to be cost effective at any feasible estimate of this cost.

Conclusions Formal identification methods for postnatal depression do not seem to represent value for money for the NHS. The major determinant of cost effectiveness seems to be the potential additional costs of managing women incorrectly diagnosed as depressed. Formal identification methods for postnatal depression do not currently satisfy the National Screening Committee's criteria for the adoption of a screening strategy as part of national health policy.

INTRODUCTION

Depression accounts for the greatest burden among all mental health problems and by 2020 is expected to become the second most common general health problem.[1] Postnatal depression is an important category of depression, with over 11% of women experiencing major or minor postnatal depression six weeks postnatally.[2] There is now considerable evidence to show that postnatal depression has a substantial impact on the mother and her partner,[3] the family,[4] mother-baby interactions,[5] and the longer term emotional and cognitive development of the baby,[6] especially when depression occurs in the first year of life.[7]

Though clinically and cost effective treatments are available,[8 9] less than half of cases of postnatal depression are detected in routine clinical practice.[9 10] Formal strategies for screening and case identification (such as standardised postnatal questionnaires, standardised generic questionnaires for depression, and prenatal screening for known risk factors for postnatal depression) have been advocated but are controversial.[11 12] The National Screening Committee has clear criteria that must be satisfied before the adoption of formal screening strategies.[13] These criteria consist of 23 items relating to the condition, the test, the treatment, and the proposed screening programme.[14] In particular, screening strategies are assessed to ensure that the screening does more good than harm at a reasonable cost. When these criteria were previously applied to formal screening strategies for postnatal depression, there was insufficient clinical and economic evidence to support their implementation.[12] Nevertheless, these strategies are widely used in current practice, with particular focus on the Edinburgh postnatal depression scale (EPDS).[8]

Furthermore, recent clinical guidelines issued by the National Institute for Health and Clinical Excellence (NICE) recommend the use of brief case finding questions to identify possible postnatal depression (box 1), with the use of self report measures such as the

Box 1: The Whooley questions

Clinical guidance issued by NICE in 2007 recommends that healthcare professionals ask two questions at a woman's first contact with primary care, again at her booking visit, and again postnatally (usually at 4-6 weeks and 3-4 months):

- 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?

Edinburgh postnatal depression scale, the hospital anxiety and depression scale (HADS), or the patient health questionnaire (PHQ-9) as part of subsequent assessment or for routine monitoring.[8] A specific recommendation was made for the use of brief generic case finding questions (the Whooley questions) that had previously been validated in older men but not postnatal women.[8] This guidance, however, did not formally consider the cost effectiveness of such strategies. In view of the uncertainty surrounding this issue, the United Kingdom National Institute for Health Research (NIHR)-Health Technology Assessment Programme prioritised a review of the clinical and cost effectiveness of formal identification methods for postnatal depression in primary care. A full technical report is published elsewhere.[15] Here we provide a summary of the cost effectiveness analysis and policy implications of this policy driven evidence review.

METHODS

We carried out several formal systematic reviews of identification methods and their corresponding performance and developed an economic model to evaluate the costs and health outcomes associated with each. This model took the form of a decision tree and considered the full treatment pathway from the possible onset of postnatal depression through identification, treatment, and possible relapse (figure).

The model evaluated a hypothetical population of women managed in primary care six weeks postnatally. We assumed that a proportion of the women were depressed but that only some had been identified as such through routine care. At this time a formal identification method could be administered on the entire population of postnatal women. As screening and case finding instruments are imperfect diagnostic instruments, this might identify some depressed women not detected through routine care, but it might also incorrectly identify some women who were not depressed. Women were assumed to have a diagnosis of depression if they were positively identified by either the formal identification method (when applicable) or routine
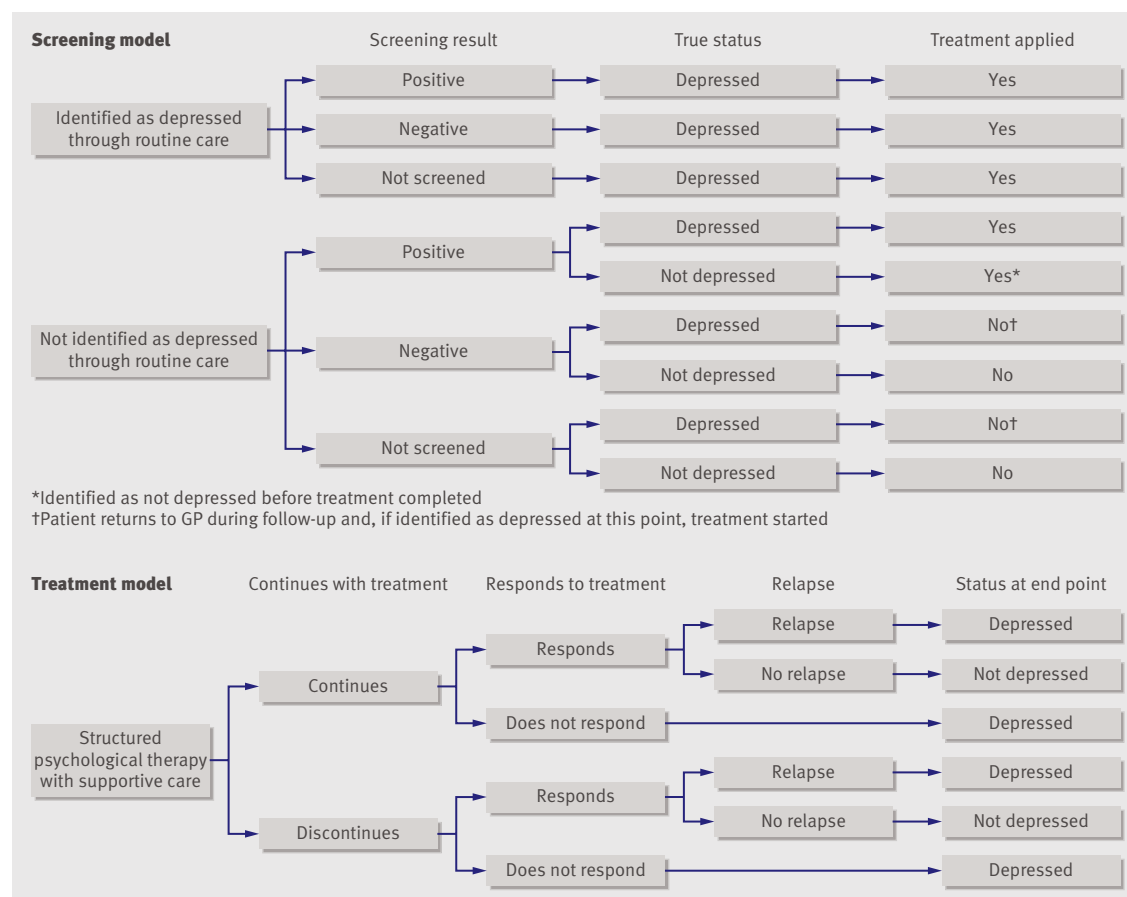
Schematic of screening and treatment models

care. We assumed that those with a diagnosis of depression were offered additional treatment (box 2) while all others received usual postnatal care. All women were followed up for a year. To facilitate the modelling of the postnatal period, we assumed that women who were not depressed six weeks postnatally remained so until the model end point.

The analysis was conducted from the perspective of the NHS and personal social services, with costs expressed in 2006-7 prices and health outcomes expressed in quality adjusted life years (QALYs). We did not apply discounting because the time horizon of the analysis was only one year. The model was probabilistic: input parameters were entered into the model as probability distributions to reflect uncertainty surrounding the mean estimates. We used Monte Carlo simulation (over 10 000 iterations) to propagate this uncertainty throughout the model and sensitivity analysis to explore the impact of alternative model assumptions.

Data sources

To specify the range of validated case finding and screening techniques and to establish their diagnostic performance characteristics, we undertook a comprehensive search across 20 electronic databases (including forward citation searching of key literature, personal communication with authors, and scrutiny of reference lists[15]). We identified numerous generic and specific screening strategies for postnatal depression, 14 of which had been validated among women during pregnancy or the postnatal period. By far the most commonly used identification method was the Edinburgh postnatal depression scale, a specially developed standardised postnatal questionnaire. Other methods identified included standardised generic questionnaires such as the Beck depression inventory. Both of these methods generate a score on a particular scale—when a patient scores at or above a particular cut point clinically significant depression is assumed to be present.[15] The performance (sensitivity and specificity) of each method depends critically on the cut point chosen (table 1).

We used a bivariate meta-analysis to derive the sensitivity and specificity of the alternative formal identification methods. This meta-analysis formed part of a wider portfolio of research that assessed the accuracy, acceptability, clinical effectiveness, and cost effectiveness of formal identification methods for postnatal depression.[15] The bivariate model was fitted with a generalised linear mixed model approach to the bivariate meta-analysis of sensitivity and specificity.[17] This approach uses the exact binomial distribution to describe variability of sensitivity and specificity within a study rather than the normal approximation as originally proposed.[18]

For "major or minor" depression there were sufficient data to pool the Edinburgh postnatal depression scale at various cut points (7-16 inclusive) and the Beck depression inventory at cut point 10. We did not consider the Whooley questions in the base case because of lack of data available to pool estimates as part of the bivariate meta-analysis and concerns as to the absence of data in postnatal women. For the purposes of the probabilistic model, we modelled the sensitivity and specificity on the log odds scale using normal distributions. The correlation between sensitivity and specificity was reflected in the probabilistic analysis using the Cholesky decomposition of the covariance matrix.[19]

We used an estimate of the prevalence of depression among postnatal women six weeks postnatally from a previous systematic review.[2] In the absence of specific data on postnatal depression, we derived the estimate of the probability that depression is detected via routine care at six weeks postnatally from a UK follow-up study of detection rates for depression and anxiety in primary care.[20] The estimate of the probability that postnatal depression is detected when a woman with undiagnosed depression returns to her general practitioner during the follow-up period was also derived from this study. The robustness of our model to uncertainty around the estimate of prospective detection rates was formally examined within our probabilistic sensitivity analysis.

Estimates of the probabilities that women discontinue treatment, do not respond to treatment, or relapse after responding to treatment (see box 2) were adopted from data used in the most recent evidence synthesis undertaken within the recent NICE clinical guideline.[8]

Table 2 summarises the model's parameters.

Box 2: Treatment for postnatal depression

The recent NICE clinical guidance presented a decision analytical model that assessed the relative cost effectiveness of two alternative psychological treatments for women with mild to moderate depression in the postnatal period: structured psychological therapy (as exemplified by manualised cognitive behavioural therapy delivered by a clinical psychologist) and non-directive counselling (for example, that based on Rogerian principles and delivered by a primary care practice counsellor); full descriptions of these treatments are given in the guidance.[8] We reconstructed this model with updated parameter inputs and carried out probabilistic Monte Carlo simulation over 10 000 iterations. By adopting a conventional threshold of willingness to pay of £20 000-30 000 per quality adjusted life year (QALY), we found that structured psychological therapy is a cost effective treatment for mild to moderate postnatal depression, with an incremental cost effectiveness ratio (ICER) of £17 481 per QALY versus no treatment. The ICER compares the additional costs that one strategy incurs over another with the corresponding additional health benefits.[16] Non-directive counselling was not cost effective, with an ICER of £66 275 per QALY versus structured psychological therapy.

Women with a diagnosis of postnatal depression in our model were therefore assumed to be offered structured psychological therapy as detailed in the NICE report. This included eight 50 minute sessions with a clinical psychologist, three 10 minute appointments with her general practitioner, four 45 minute home visits from a health visitor (a qualified nurse specially trained in health assessment and promotion in the community), and a single one hour home visit from a community psychiatric nurse. As in the NICE model, each woman would either continue or discontinue treatment and would either respond to treatment or not. If the woman responded then there was a possibility of relapse. If the woman did not respond then she remained depressed.

Women with an incorrect diagnosis were assumed not to receive any of the sessions with the clinical psychologist but did receive the supportive care. Women with undiagnosed depression were assumed to return to their general practitioner during the follow-up period, at which time there was a possibility of receiving a diagnosis and being offered treatment; meanwhile there was a possibility of these women recovering under usual postnatal care. Full details are provided in the technical report.[15]

Table 1 | Performance of two methods of identification of postnatal depression in base case according to different cut points

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

Quality of life

Critical to the cost effectiveness of screening and case finding for postnatal depression is the impact of early identification and management on a woman's quality of life. To estimate QALYs, we assessed the quality of life associated with different health states in terms of preference weights based on published evidence.[21] In the absence of published weights specific for postnatal depression, we assumed that affected women experience the quality of life associated with "moderate depression," while women without depression (including those in remission) experience the quality of life associated with "remission."

Costs

The costs associated with formal identification methods are also critical to their cost effectiveness. These include the cost of administering the method, the cost of any subsequent treatment, and the costs associated with incorrect diagnosis. We established estimates of time and cost from published data or from expert opinion.

Estimates for NHS unit costs were derived from national reference costs.[22] In common with the assumptions used with the NICE clinical guideline, structured psychological therapy consisted of eight 50 minute sessions with a clinical psychologist (£446.67) plus the following supportive care: three 10 minute appointments with a general practitioner (£76.50), four 45 minute home visits from a health visitor (£278.20), and a single one hour home visit from a community psychiatric nurse (£59.30).

Estimated administration costs were £7.57 (five minutes of a health visitor's time at £91 an hour) for the Edinburgh postnatal depression scale and £8.59 (five minutes at £91 an hour plus the $2 licence fee) for the Beck depression inventory.[23]

Women with undiagnosed depression were assumed to make an additional 10 minute visit to their general practitioner during the follow-up period (£25.50). We assumed that women without depression who were wrongly identified as depressed would not receive any of the sessions with the clinical psychologist but would receive all the supportive care (total cost £414); this assumption was returned to in a sensitivity analysis.

Sensitivity analyses

We identified several important drivers of the model: the cost incurred by and approach to managing those wrongly identified as depressed (false positives); and the decision to include or exclude formal identification strategies not evaluated in the bivariate meta-analysis because of insufficient data. The impact of these drivers on the results of the model was explored through sensitivity analysis with a series of alternative scenarios.

Scenario 1—The cost incurred by managing women with a false positive diagnosis is a particularly important driver of the model that is subject to considerable uncertainty. The base case estimate of this cost (£414) was calculated on the assumption that women wrongly identified as depressed would receive all of the supportive care associated with structured psychological therapy before being identified as not depressed. As this estimate might well be conservative, we considered an alternative scenario whereby this cost was assumed to be that of a single consultation with a general practitioner (£25.50), on the optimistic assumption that women identified as depressed by a formal identification method would be referred to a general practitioner, who would then immediately make the correct diagnosis.

Scenario 2—We also considered the impact of using a complementary ideal identification test, such as the structured clinical interview for DSM-IV axis I disorders (SCID), for those identified as depressed by a formal identification method. We assumed that this interview took 30 minutes of a health visitor's time to administer (£45.50) and had a sensitivity and specificity of 100%; this mitigated the cost of managing false positives.

Scenario 3—We evaluated the impact of considering the Whooley questions as an alternative identification

Table 2 | Parameters of model used to determine cost effectiveness of screening for postnatal depression in primary care

Parameter	Mean (SE*)	Distribution	Source
Clinical parameters			
Prevalence of major or minor depression	0.1130 (0.0221)	Normal	Gaynes et al (2005)[2]
Probability of identification of depression via routine care at 6 weeks postnatally	0.3864 (0.0516)	Beta	Kessler et al (2002)[20]
Probability of GP identification of depression at 32 weeks postnatally	0.4074 (0.0663)	Beta	Kessler et al (2002)[20]
Absolute risk of discontinuation (routine care)	0.0861 (0.0287)	Normal	NICE (2007)[8] (revised)
Relative risk of discontinuation (structured psychological therapy)	2.66 (1.9783)	Normal	NICE (2007)[8] (revised)
Absolute risk of no improvement (routine care)	0.6037 (0.0514)	Normal	NICE (2007)[8] (revised)
Relative risk of no response (structured psychological therapy)	0.63 (0.1276)	Normal	NICE (2007)[8] (revised)
Absolute risk of relapse (routine care)	0.3120 (0.0752)	Normal	NICE (2007)[8] (revised)
Relative risk of relapse (structured psychological therapy)	0.59 (0.2525)	Normal	NICE (2007)[8] (revised)
Utility parameters			
Moderate depression	0.63 (0.0275)	Beta	Revicki and Wood (1998)[21]
Remission without maintenance treatment	0.86 (0.0191)	Beta	Revicki and Wood (1998)[21]
Cost parameters			
BDI licence fee (per test)	$2/£1.02†	Fixed	Pearson[23]; FT.com
Health visitor (per hour)	£91.00	Fixed	Curtis (2007)[22]
Health visitor (travel)	£1.30	Fixed	Curtis (2007)[22]
Clinical psychologist (per hour)	£67.00	Fixed	Curtis (2007)[22]
General practitioner (per hour)	£153.00	Fixed	Curtis (2007)[22]
Community psychiatric nurse (per hour)	£58.00	Fixed	Curtis (2007)[22]
Community psychiatric nurse (travel)	£1.30	Fixed	Curtis (2007)[22]
EPDS (health visitor 5 minutes)	£7.57	Fixed	Estimated from above inputs
BDI (health visitor 5 minutes + license fee)	£8.59	Fixed	Estimated from above inputs
SCID (health visitor 30 minutes)	£45.50	Fixed	Estimated from above inputs
Structured psychological therapy	£446.67	Fixed	Estimated from above inputs
Supportive care	£414.00	Fixed	Estimated from above inputs

EPDS=Edinburgh postnatal depression scale; BDI=Beck depression inventory; SCID=structured clinical interview for DSM-IV.
*Where applicable.
†At 1 April 2008.

method, given its recommendation in the recent NICE clinical guidance and its subsequent policy relevance. This was not considered in the base case because of the lack of data available to pool estimates as part of the bivariate meta-analysis and concerns as to the absence of validation data in a postnatal population. Given these limited data, the Whooley questions were assumed to have the sensitivity (0.96, 95% confidence interval 0.86 to 0.99) and specificity (0.89, 0.87 to 0.91) reported by Arroll et al in 2005[24] and were assumed to take a health visitor one minute to administer (£1.52).

RESULTS

We have presented our results in two ways: firstly, mean lifetime costs and QALYs of each formal identification method with the cost effectiveness of each method compared with incremental cost effectiveness ratios (see box 2); and, secondly, decision uncertainty as the probability that each formal identification method is considered the most cost effective strategy for a given cost effectiveness threshold.

Table 3 summarises the results of the base case analysis. No formal identification method seemed to be cost effective under a conventional willingness to pay threshold of £20 000-30 000 per QALY,[25] with the strategy of routine care the most likely to be cost effective. Adopting the Edinburgh postnatal depression

scale with a cut point of 16 was the least costly and effective formal identification method, with an estimated incremental cost effectiveness ratio of £41 103 per additional QALY compared with routine care alone. The incremental cost effectiveness ratio for all other screening strategies ranged from £49 928 to £272 463 per QALY versus routine care only. The costs of managing those wrongly diagnosed as depressed represented a key driver of these results; as such, strategies adopting formal identification methods with higher specificity were associated with more favourable incremental cost effectiveness ratios.

Sensitivity analyses

Table 4 summarises the results of the sensitivity analyses. The assumption of a lower cost associated with managing those wrongly diagnosed as depressed had two important consequences: the incremental cost effectiveness ratios of strategies adopting identification methods became more favourable; and the ranking of non-dominated strategies became less dependent on the specificity of the identification method adopted. When the cost of managing those wrongly diagnosed was assumed to be £25.50, using the Edinburgh postnatal depression scale with a cut point of 10 seemed borderline cost effective with an incremental cost effectiveness ratio of £29 186 per additional QALY.

Table 3 | Results of base case analysis of screening for postnatal depression in primary care

	Mean QALYs	Mean costs	ICER†	Probability cost effective*	
				£20 000	£30 000
Routine care only	0.8455	£49.29	NA	87.65	58.69
EPDS cut point:					
16	0.8461	£73.49	£41 103	2.21	6.14
15	0.8462	£80.95	ED	0.70	1.82
14	0.8465	£94.21	£49 928	1.58	4.39
13	0.8467	£110.47	D	0.52	2.53
12	0.8468	£109.95	£56 697	1.77	6.11
BDI cut point:					
10	0.8468	£121.51	D	1.15	5.07
EPDS cut point:					
11	0.8468	£118.82	£113 411	1.86	5.87
10	0.8470	£140.44	£120 968	1.72	5.64
9	0.8471	£156.95	£245 210	0.65	2.60
7	0.8472	£215.07	D	0.01	0.04
8	0.8472	£187.32	£272 463	0.18	1.10

ICER=incremental cost effectiveness ratio; NA=not applicable; EPDS=Edinburgh postnatal depression scale; BDI=Beck depression inventory.
*Probability (in percentage) that strategy is more cost effective than all others given particular maximum willingness to pay for additional QALY.
†D=dominated strategy (one or more alternative strategies are cheaper and more effective); ED=extended dominance (one or more alternative strategies are more expensive but have lower ICER).

Under base case assumptions, adoption of the structured clinical interview as a confirmatory test proved to be cost saving compared with the equivalent strategy without the interview, although no strategy subsequently proved to be cost effective based on a threshold of £20-30 000 per QALY. Use of the Edinburgh postnatal depression scale with a cut point of 13 with confirmatory structured clinical interview had the lowest reported incremental cost effectiveness ratio of £33 776 per QALY compared with routine care.

When we considered the Whooley questions as an alternative identification method they proved not to be cost effective, with an incremental cost effectiveness ratio of £46 538 per QALY versus the strategy of using the Edinburgh postnatal depression scale with a cut point of 16 (itself not cost effective compared with routine care only).

DISCUSSION

The use of formal identification methods for detecting postnatal depression does not to represent value for money for the NHS. Postnatal depression is an important clinical, economic, and social problem which is under-recognised and for which effective treatments are available. Decisions to screen, however, have attracted considerable controversy, and such policy decisions should be informed by systematic consideration of the clinical and economic evidence. In the absence of prospective economic evidence collected within randomised trials, decision models remain the most useful method to inform practice and policy decisions and to identify areas of uncertainty and priorities for further research. We used decision modelling to address this problem and built on a recent and comprehensive diagnostic meta-analysis of the performance of screening instruments.

Our conclusions regarding the lack of cost effectiveness are primarily driven by the costs of managing false positives—that is, women with a misdiagnosis of depression at a one off screen who do not subsequently turn out to have postnatal depression. When this cost is relatively high, the specificity of a formal identification method is an important contributor to its cost effectiveness; as this cost falls, specificity becomes less important relative to sensitivity. In the absence of reliable data surrounding this cost, we used a conservative approach in our base case analysis: compared with routine care the most cost effective formal identification method (Edinburgh postnatal depression scale with a cut point of 16) had an incremental cost effectiveness ratio of £41 103 per QALY. This ratio is well above a conventional threshold of willingness to pay of £20 000 -30 000 per QALY. Furthermore, even when we adopted a particularly optimistic estimate of this cost (that of a single appointment with a general practitioner) in a sensitivity analysis, compared with routine care the most cost effective formal identification method (Edinburgh postnatal depression scale with a cut point of 10) had an incremental cost effectiveness ratio of £29 186 per QALY, falling only slightly under the upper limit of the £20 000-30 000 per QALY cost effectiveness threshold. As such, our main finding seems to be robust to plausible estimates of the cost of managing false positives. Nevertheless, a definitive answer as to the cost effectiveness of formal identification methods requires further evidence around this particular cost.

A secondary finding is that adopting the structured clinical interview as a confirmatory test for those positively identified by a formal identification method proved to be cost saving compared with the equivalent strategy without such an interview; however, no such strategy subsequently proved to be cost effective compared with routine care only. This suggests that future research into the cost associated with managing false positives should also consider alternative approaches to the management of those positively identified by a formal identification method.

Strengths and limitations

We evaluated the relative performance of each identification method using bivariate meta-analysis, with the data used to inform the meta-analysis derived from several systematic searches of the literature. We used probabilistic techniques to propagate parameter uncertainty throughout the model. The model is compatible with the most recent best practice guidelines from NICE for the methods of technology appraisal.[25]

The analysis was conducted from the perspective of the NHS and personal social services and the model focused on the costs and health outcomes associated solely with the mother; no account was taken of the potential impact that successful identification and subsequent management might have had on other family

Table 4 | Results of sensitivity analyses of screening for postnatal depression in primary care results

| | Mean QALYs | Mean costs | ICER | Probability cost effective* | |
				£20 000	£30 000
Scenario 1: Managing patients wrongly identified as depressed costs £25.50					
Routine care only	0.8452	£50.33	NA	74.66	43.59
EPDS cut point:					
10	0.8468	£95.26	£29 186	5.02	9.34
8	0.8469	£100.74	£35 390	5.78	14.95
Scenario 2: SCID is adopted as confirmatory test					
Routine care only	0.8450	£50.11	NA	83.62	52.61
EPDS cut point:					
13	0.8462	£91.30	£33 776	1.14	3.16
10	0.8465	£102.49	£37 391	2.96	8.29
8	0.8466	£110.00	£50 408	2.26	8.58
Scenario 3: Whooley questions is a feasible strategy					
Routine care only	0.8455	£49.34	NA	88.52	59.78
EPDS cut point:					
16	0.8461	£73.54	£41 175	2.08	5.15
Whooley questions	0.8473	£130.16	£46 538	0.27	3.43

ICER=incremental cost effectiveness ratio; EPDS=Edinburgh postnatal depression scale; NA=not applicable; SCID=structured clinical interview for DSM-IV.

*Probability (in percentage) that strategy is more cost effective than all others given particular maximum willingness to pay for additional QALY for non-dominated or extendedly dominated strategies only.

members or the infant(s). These aspects, and a wider societal perspective, were not considered because of the lack of reliable evidence on the wider impact of case identification or treatment strategies.

There were limited published data available for estimating particular parameters in the model, including the probability that postnatal depression was identified via routine care at six weeks, the risk of relapse, and the utility weights. As a result, we derived the estimates used in the model from studies of general depressed populations (that is, not postnatal women), which represents a serious limitation of the model, particularly with respect to the utility estimates adopted. Further research specifically into the health related quality of life of women with postnatal depression would be valuable for future studies. The Edinburgh postnatal depression scale 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; as such, the performance of other identification strategies could not be assessed at multiple cut points. In addition, there were insufficient data to carry out subgroup analyses on different populations, such as women of different ages and women with or without complications after birth.

A further issue is the degree to which the QALY is an appropriate measure of health outcome. While the QALY is used throughout the literature on evaluation of health economics, it might be an insensitive measure of outcomes in mental health care.[26] In the absence of a suitable alternative, we used the QALY to ensure comparability between the interventions considered here and those outside mental health; the potential insensitivity of the QALY in this context, however, should be considered in the interpretation of the results.

Finally, there were moderate to high levels of heterogeneity between studies across all cut points of the instruments considered. Consequently, we used random effects meta-analysis to incorporate the additional uncertainty caused by that heterogeneity in the test performance results for each instrument. It has recently been argued that conventional random effect approaches might underestimate uncertainty by not considering the whole distribution of effects and that this additional uncertainty can be reflected by using a prediction interval.[27] In the context of decision models analysts could use either the predictive distribution of a future treatment effect or they should assume that future implementation will result in a distribution of treatment effects.[28] The importance of this issue depends on the sources of heterogeneity in the evidence and their relation to a future implementation, both of which remain uncertain here given the limited existing evidence base for several of the instruments. Importantly from a policy perspective, the decision model presented here is linear with respect to the parameter inputs and the outputs used to establish the incremental cost effectiveness ratios and hence our conclusions would not be altered by using the predictive distribution.

Policy issues

The results suggest that application of the recent NICE guidance (which recommends the use of the Whooley questions with an additional help question, see box 1) and widespread current practice (which focuses on the routine or ad hoc administration of the Edinburgh postnatal depression scale) do not result in value for money for the NHS. Formal methods of screening or case identification do not seem to satisfy the National Screening Committee's criteria for the adoption of a

WHAT IS ALREADY KNOWN ON THIS TOPIC

Though cost effective treatments for postnatal depression are available, fewer than half of cases of postnatal depression are detected by primary healthcare professionals in routine clinical practice

Screening and case identification strategies have been advocated but have attracted substantial controversy

Recent NICE guidance recommended the use of brief case finding questions (the Whooley questions) to identify possible postnatal depression

WHAT THIS STUDY ADDS

Routine screening for postnatal depression does not seem to represent value for money for the NHS, even with the use of a confirmatory test for those identified as depressed and with adjustment for the costs of managing those misdiagnosed

Routine screening for postnatal depression does not satisfy the National Screening Committee's criteria for the adoption of a screening strategy as part of national health policy

screening strategy as part of national health policy as their adoption does not represent a "favourable ratio of costs to benefits" according to conventional NHS cost effectiveness thresholds.

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. Harvard School of Public Health on behalf of the World Bank, 1996.

2 Gaynes BN, Gavin N, Meltzer-Brody S, Lohr K, Swinson T, Gartlehner G, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. 2005. www.ahrq.gov/downloads/pub/evidence/pdf/peridepr/peridep.pdf.

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

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

5 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.

6 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.

7 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.

8 NICE. Antenatal and postnatal mental health: clinical management and service guidance. 2007. www.nice.org.uk/CG45.

9 Hendrick V. Treatment of postnatal depression. *BMJ* 2003;327:1003-4.

10 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.

11 Henshaw C, Elliott S. Screening for perinatal depression. Jessica Kingsley Publishers, 2005.

12 Shakespeare J. Screening: the role and recommendations of the UK National Screening Committee. In: Henshaw C, Elliott S, eds. Screening for perinatal depression. Jessica Kingsley Publishers, 2005:21-33.

13 National Screening Committee. The UK National Screening Committee's criteria for appraising the viability, effectiveness and appropriateness of a screening programme. Stationery Office, 2003.

14 UK National Screening Committee. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. 2009. www.screening.nhs.uk/criteria.

15 Hewitt C, Gilbody S, Brealey S, Paulden M, Palmer S, Mann R, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. *Health Technol Assess* 2009;13:1-145, 147-230.

16 Johannesson M. On the estimation of cost-effectiveness ratios. *Health Policy* 1995;31:225-9.

17 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.

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

19 Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation. Oxford University Press, 2006.

20 Kessler D, Bennewith O, Lewis G, Sharp D. Detection of depression and anxiety in primary care: follow up study. *BMJ* 2002;325:1016-7.

21 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.

22 Curtis K. Unit costs of health and social care 2006. Personal Social Services Research Unit, University of Kent, 2007.

23 Pearson. Beck depression inventory-II (BDI®-II). www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8018-370&Mode=summary.

24 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.

25 NICE. Guide to the methods of technology appraisal. 2008. http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf.

26 Chisholm D, Healey A, Knapp M. QALYs and mental health care. *Soc Psychiatry Psychiatr Epidemiol* 1997;32:68-75.

27 Higgins J, Thompson S, Spiegelhalter D. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;172:137-59.

28 Ades AE, Lu G, Higgins JPT. The interpretation of random-effects meta-analysis in decision models. *Med Decis Making* 2005;25:646-54.

Accepted: 8 October 2009