Psychological Medicine

cambridge.org/psm

Original Article

Cite this article: Henderson C, Dixon S, Bauer A, Knapp M, Jane Morrell C, Slade P, Walters SJ, Brugha T (2018). Cost-effectiveness of PoNDER health visitor training for mothers at lower risk of depression: findings on prevention of postnatal depression from a cluster-randomised controlled trial. *Psychological Medicine* 1–11. https://doi.org/10.1017/S0033291718001940

Received: 17 November 2017
Revised: 27 June 2018
Accepted: 2 July 2018

Key words:
Cost-effectiveness; depression; health visitor; perinatal; prevention

Author for correspondence:
Catherine Henderson, E-mail: c.henderson@lse.ac.uk

© Cambridge University Press 2018. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

CAMBRIDGE UNIVERSITY PRESS

Cost-effectiveness of PoNDER health visitor training for mothers at lower risk of depression: findings on prevention of postnatal depression from a cluster-randomised controlled trial

Catherine Henderson[1], Simon Dixon[2], Annette Bauer[1], Martin Knapp[1], C. Jane Morrell[3], Pauline Slade[4], Stephen J. Walters[2] and Traolach Brugha[5]

[1]Personal Social Services Unit, London School of Economics and Political Science, London, UK; [2]School of Health and Related Research, University of Sheffield, Sheffield, UK; [3]School of Health Sciences, University of Nottingham, Nottingham, UK; [4]Institute of Psychology Health and Society, University of Liverpool, Liverpool, UK and [5]Department of Health Sciences, University of Leicester, Leicester, UK

Abstract

Background. There is evidence for the cost-effectiveness of health visitor (HV) training to assess postnatal depression (PND) and deliver psychological approaches to women at risk of depression. Whether this approach is cost-effective for lower-risk women is unknown. There is a need to know the cost of HV-delivered universal provision, and how much it might cost to improve health-related quality of life for postnatal women. A sub-study of a cluster-randomised controlled trial in the former Trent region (England) previously investigated the effectiveness of PoNDER HV training in mothers at lower risk of PND. We conducted a parallel cost-effectiveness analysis at 6-months postnatal for all mothers with lower-risk status attributed to an Edinburgh Postnatal Depression Scale (EPDS) score <12 at 6-weeks postnatal.

Methods. Intervention HVs were trained in assessment and cognitive behavioural or person-centred psychological support techniques to prevent depression. Outcomes examined: quality-adjusted life-year (QALY) gains over the period between 6 weeks and 6 months derived from SF-6D (from SF-36); risk-of-depression at 6 months (dichotomising 6-month EPDS scores into lower risk (<12) and at-risk (⩾12).

Results. In lower-risk women, 1474 intervention (63 clusters) and 767 control participants (37 clusters) had valid 6-week and 6-month EPDS scores. Costs and outcomes data were available for 1459 participants. 6-month adjusted costs were £82 lower in intervention than control groups, with 0.002 additional QALY gained. The probability of cost-effectiveness at £20 000 was very high (99%).

Conclusions. PoNDER HV training was highly cost-effective in preventing symptoms of PND in a population of lower-risk women and cost-reducing over 6 months.

Introduction

One in five women may experience perinatal mental illness (Gavin *et al.*, 2005). Opportunities to diagnose these conditions may be missed by primary care physicians (Prady *et al.*, 2016; Ford *et al.*, 2017) and, in particular, half of all cases of perinatal depression (prevalence of one in eight women) go undetected (Gavin *et al.*, 2015; Bauer *et al.*, 2016*b*). Consequences of untreated postnatal depression (PND) can be profound and long-lasting for women and families, with risks of longer-term adverse effects on child development and associated costs (Bauer *et al.*, 2016*b*), yet there are psychosocial treatments that a growing evidence base suggests are effective (Milgrom and Gemmill, 2014; Morrell *et al.*, 2016).

Health visitors (HVs) are public health nurses, based in community settings such as health centres and family centres across the UK, playing important roles in supporting women during and after pregnancy (Cowley *et al.*, 2007, 2013; Health Education England, 2016). The impact of PoNDER HV training in assessment and delivery of cognitive-behavioural and person-centred approaches (CBA and PCA) in terms of effectiveness and cost-effectiveness for women and children has been reported (Morrell *et al.*, 2009*a*). PoNDER HV training can reduce the proportion of women at risk of developing PND as indicated by a reduction in score on the Edinburgh Postnatal Depression Scale (EPDS) (Cox *et al.*, 1987), a self-report measure widely used in clinical practice (Hewitt and Gilbody, 2009). Scores on the 10-item EPDS range from 0 to 30, higher scores indicating more depressive symptoms; at-risk women were identified as those scoring 12 or more.

Further analyses of mental health outcomes in the subgroup of *lower-risk* women (EPDS score < 12) in the PoNDER trial at 6 weeks after childbirth suggested that participants in

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

the intervention group had a reduced risk of developing PND as indicated by reduction in EPDS scores 6 months after childbirth (Brugha *et al.*, 2011). While the training intervention appeared to be effective, that analysis did not address the separate question of costs and cost-effectiveness. There is a need to know whether the additional costs associated with providing this additional care are considered worthwhile in relation to the health benefit it produces. Here we examine the cost-effectiveness of this *universal preventive approach* for lower-risk women.

Method

The main PoNDER cluster-randomised controlled trial is detailed elsewhere (Morrell *et al.*, 2009*b*, 2009*a*). The trial randomised 101 general practitioner (GP) practices to either (a) 'usual health visitor care' ($n = 38$ clusters; 1 cluster lost to follow-up); (b) care by HVs trained in assessing postnatal for symptoms of PND and a CBA to address postnatal psychological problems ($n = 30$ clusters); or (c) care by HVs trained in assessing women for symptoms of PND and a PCA to address postnatal psychological problems ($n = 32$ clusters). HVs received brief training derived from CBT principles (Appleby *et al.*, 1997) for the CBA arm or from person-centred counselling principles (Holden *et al.*, 1989), for the PCA arm (Morrell *et al.*, 2011). The training was carried out by Masters-level trainers and was equivalent for both the CBA and PCA arms (1 day on clinical assessment skills; 5 days on psychotherapeutic approach; 4 half days of reflective practice/clinical supervision). The training developed intervention HVs' skills in genuineness and listening so that HVs could talk to the woman about PND, gain her trust, develop an ongoing relationship with her, and be open to being re-contacted if the woman felt that her symptoms were not improving spontaneously. We recognised that the research was undertaken in a population that could be considered to be a vulnerable group and so the trial excluded: women under 18 and women with pre-existing severe and enduring mental health problems. The trial began in April 2003 and continued over 3 years. Baseline measurements, including EPDS, were taken at 6 weeks postnatally. Women completed study questionnaires at 6, 12, and 18 months postnatally. In all, 4084 women consented to participate in the study; 3449 returned baseline 6-week postal questionnaires, of whom 595 (17.3%) scored 12 or more on EPDS ('at-risk' women). At-risk women in intervention groups (b) and (c) were offered up to 8 weekly psychologically informed sessions with the HV. Brugha *et al.* (2011) subsequently reported a sub-study of the main trial, focused on lower-risk participants who scored *below* 12 on the 6-week postnatal EPDS ('EPDS-negative') (further information, including trial CONSORT diagram and cluster-randomisation methods can be found in that publication). The study examined the effectiveness of care from HVs trained in assessment and psychological support in preventing PND in EPDS-negative women 6–18 months later.

Economic evaluation

The economic evaluation examined the cost-effectiveness of PoNDER HV training for PND in the population of lower-risk (EPDS-negative) women examined in the Brugha *et al.* (2011) sub-study. We also explored whether PoNDER training affected the number of HV visits and whether the impact was similar across women with different levels of depression risk. The economic evaluation design followed technology appraisal guidelines by the National Institute for Clinical Excellence (2004) (NICE) (now National Institute for Health and Care Excellence) and consequently took an NHS and social care perspective.

Costs

The following costs were included: HV training and ongoing clinical supervision to deliver the intervention; HV contacts (number, duration and purpose of visit – whether for PND, mother (excluding PND), baby or any combination of the three); infant immunisations; GP contacts; prescriptions for all conditions; social worker contacts; admissions to Mother and Baby psychiatric units; other mental health contacts, including counsellor, community psychiatric nurse (CPN), community mental health team (CMHT); mental health nurse (MHN), crisis services, psychologist, psychotherapist, psychiatric outpatient and mother and baby psychiatric outpatient.

Cost components (Table 1) other than costs of delivering the intervention were derived from resource use data and nationally representative unit costs applicable at the time the trial started [principally Netten and Curtis (2004)]. Resource use data from 6 weeks to 6 months were collected on a resource use log completed by HVs based on their own and GP records (Morrell *et al.*, 2009*b*).

The costs of delivering PoNDER training were calculated from trainer fees, travel, backfill HV time and ongoing clinical supervision costs (Morrell *et al.*, 2009*b*). Mean cost of training per HV was £1398 (annual equivalent £988). This translated to an increase in cost per HV hour of client time from £77 (Netten and Curtis, 2004) to £79: these figures represent unit costs used for HVs in control and intervention groups, respectively.

Outcomes

Data on health-related quality of life were collected using SF-36 at 6 weeks, 6 and 12 months, and generated the preference-based health measure, the SF-6D index (Brazier *et al.*, 2002). Mothers' quality-adjusted life-year (QALY) gains between 6 weeks and 6 months were calculated from SF-6D utility index scores at those time points by the area-under-the-curve using the trapezoidal method (Manca *et al.*, 2005*a*). We examined risk-of-depression outcomes at follow-up as a secondary outcome, dichotomising 6-month EPDS scores into 1 = lower-risk (score < 12) and 0 = at-risk (score ⩾ 12).

Analysis

Cluster-adjusted *t* tests were applied to comparisons of continuous data; cluster-adjusted chi-squared tests were applied to tabulations of categorical data (Davis, 2001; Herrin, 2012). Intra-cluster correlation coefficients (ICCs) were derived from one-way analysis of variance (Ukoumunne, 2002). The conventional 5% level of significance was used throughout.

In addition to cost-effectiveness analyses, we explored whether the PoNDER HV training affected the number of HV visits and whether the impact varied by level of depression risk. We addressed the latter question by examining whether numbers of HV visits over the 6-month follow-up differed according to the proximity of participants' scores to the threshold for '*lower-risk*' *v.* '*at-risk women*'. The sample was differentiated into risk subgroups of 'very low risk', 'subthreshold risk' and 'at-risk' women (EPDS score 0–5; 6–11; and 12 or more, respectively) adopting

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

Table 1. Unit costs of resources used

Resource	Unit cost (£s 2003/4)	Source
HV hour of client contact without CBA/PCA training	77	Netten and Curtis (2004)
HV hour of client contact with CBA/PCA training	79	Study records (and using HV contacts data from Table 2)
GP contact[a]	30	Netten and Curtis (2004)
Social work visit[b]	108	Netten and Curtis (2004)
Clinical mental health contact[c,d]	129	Trust Financial Returns 2004 (Department of Health, 2000)
Community mental health contact[e]	29	Netten and Curtis (2004)
Mother and baby or psychiatric unit day[c]	458	Reference costs 2004 (Department of Health, 2005)
Fluoxetine prescription[f]	2	British National Formulary 2005 (British Medical Association and Royal Pharmaceutical Society, 2011)
Other prescription[g]	3	British National Formulary 2005 (British Medical Association and Royal Pharmaceutical Society, 2011)
DTwP and Hib vaccination per dose	20	British National Formulary 2005 (British Medical Association and Royal Pharmaceutical Society, 2011)
Men-C vaccination per dose	18	British National Formulary 2005 (British Medical Association and Royal Pharmaceutical Society, 2011)
Inpatient admission (infant)[c]	516	Trust Financial Returns 2004 (Department of Health, 2000)
A&E attendance[c]	73	Reference costs 2004 (Department of Health, 2005)
NHS direct contact	25	Hansard (Hansard, 2004*a*, 2004*b*)
Walk-in centre attendance[c]	39	Reference costs 2004 (Department of Health, 2005)

[a]Includes surgery, home and telephone contacts. Unit cost based on the most common type of contact; surgery contact.
[b]Assuming a 2-hour visit. No information was available on length of visit.
[c]Prices adjusted using inflation indices given in Netten and Curtis, 2004. (Netten and Curtis, 2004).
[d]Includes crisis service, psychologist, psychotherapist, psychiatric outpatient and mother and baby psychiatric outpatient contacts. Unit cost based on the most common type of contact; psychiatric outpatient contact.
[e]Includes counsellor, community psychiatric nurse (CPN), community mental health team (CMHT) and mental health nurse (MHN) contacts. Unit cost based on the most common type of contact; CPN home visit.
[f]Based on most common drug and dosage for antidepressant prescriptions.
[g]Calculated as an average of the prescriptions for the eight most common indications.

cut-points used previously (Brugha *et al.*, 2011) for the effectiveness analysis. A two-level negative binomial model of HV visits was fitted to examine impacts of PoNDER training on visits across all low-risk participants, controlling for experimental group, reason for visit (for mother, baby and/or PND), number of children, history of serious life-events and (for reasons discussed below) cluster size and treatment × cluster size interaction. The model was then extended to include risk sub-group (very low risk *v.* sub-threshold risk) and its interaction with the experimental group.

Analysis of cluster-randomised data must consider correlations between observations within clusters to avoid biased estimates of sampling uncertainty and imprecision in estimating coefficients (Manca *et al.*, 2005*b*). HVs and GPs working within the general practice and the women registered there comprised a cluster. Clustering effects can arise because a practice's patients might have characteristics in common (e.g. similar reasons for living within the practice area, socio-economic circumstances).

Clustering may affect costs and outcomes differently, and ICCs of costs may be larger than those of outcomes (Gomes *et al.*, 2012*a*). There may be differential recruitment to and attrition from clusters between trial arms (Adams *et al.*, 2004). Imbalances in cluster size between experimental groups could be related to the outcome being measured; for example, the volume of work undertaken by practitioners may be related to patient outcomes (Panageas *et al.*, 2007; Gomes *et al.*, 2012*a*).

Because of its cluster-randomised design, the PoNDER trial dataset (*all* levels of risk of PND) has previously served as the basis for exploring approaches to economic modelling of clustered data (Gomes *et al.*, 2012*b*). We applied recommended methods (Gomes *et al.*, 2012*b*) to all lower-risk participants (EPDS score<12), considering effects of clustering on estimated coefficients and standard errors of costs and outcomes while addressing potential correlations between them (Gomes *et al.*, 2012*a*). In our base-case analysis, we used a system of equations (seemingly-unrelated regressions, SUR) where costs and outcomes error terms are permitted to be correlated. The system yields a coefficient on the treatment allocation term in both equations to enable estimation of cost/outcome differences between groups and the covariance between those coefficients.

To adjust for imbalances in cluster size, we incorporated a treatment × cluster size interaction term into each equation (Gomes *et al.*, 2012*b*). Cost and outcome (QALY; dichotomous 6-month depression-risk) equations adjusted for potential confounders: mother's age, history of PND, living alone, any history of major life-events, baseline 6-week EPDS score, number of other children in the family and whether mother was economically active. In an analysis of the primary outcome, costs and QALY equations additionally controlled for 6-week (baseline) utility (Manca *et al.*, 2005*a*). The impact of clustering on estimation precision was accounted for by calculating cluster-robust errors.

Incremental cost-effectiveness ratios (ICERs) and cost-effectiveness acceptability curves (CEACs) were generated from cost and outcome regressions. CEACs show the probability that an intervention is cost-effective at various hypothetical 'threshold values' of an outcome. The range of willingness-to-pay values covered by the CEAC included £20 000 per QALY, the lower range of thresholds typically used by NICE to identify which interventions to recommend for implementation (National Institute for Clinical Excellence, 2004; National Institute for Health and Clinical Excellence, 2008; National Institute for Health and Care Excellence, 2013). While women with complete clinical and economic data (complete cases) were considered in the base-case analysis, imputation of 6-month missing data was carried out in sensitivity analyses.

Sensitivity analyses

To explore whether primary outcome (QALY) findings were robust to assumptions in important parameters, we varied the

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

Table 2. Resource use at 6-months, available cases

Item	Control Mean (s.e.) *n* = 417	Intervention (combined) Mean (s.e.) *n* = 1042	Intervention (CBA) Mean (s.e.) *n* = 553	Intervention (PCA) Mean (s.e.) *n* = 489	Intervention (combined) Mean difference from control (95% confidence interval)	Intervention (CBA) Mean difference from control (95% confidence interval)	Intervention (PCA) Mean difference from control (95% confidence interval)
HV total contacts[a]	8.0 (0.8)	6.4 (0.5)	6.7 (0.8)	6.1 (0.9)	−1.6 (−3.5, 0.3)	−1.3 (−3.7, 1.1)	−1.9 (−4.4, 0.7)
HV contacts for baby[a]	7.2 (0.7)	6.1 (0.5)	6.4 (0.7)	5.7 (0.8)	−1.2 (−2.9, 0.5)	−0.8 (−2.9, 1.3)	−1.6 (−3.7, 0.6)
HV contacts for mother[a]	3.6 (0.5)	2.0 (0.3)	1.4 (0.4)	2.7 (0.5)	−1.6 (−2.7, −0.4)**	−2.1 (−3.5, −0.8)**	−0.9 (−2.3, 0.5)
HV contacts for PND[a]	0.3 (0.1)	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	−0.1 (−0.3, 0.1)	−0.0 (−0.3, 0.2)	−0.1 (−0.3, 0.1)
Total HV minutes	172.7 (20.7)	116.8 (13.6)	110.7 (20.3)	123.7 (23.8)	−55.9 (−105.4, −6.5)*	−62.1 (−123.9, −0.2)*	−49.0 (−117.4, 19.4)
GP contacts	2.5 (0.2)	2.3 (0.1)	2.3 (0.2)	2.3 (0.2)	−0.2 (−0.7, 0.3)	−0.2 (−0.8, 0.4)	−0.1 (−0.7, 0.5)
Mother and baby unit days	0	0	0	0	0	0	0
Community mental health contacts[b]	0	0	0	0	0	0	0
Clinical mental health contacts[c]	0	0	0	0	0	0	0
A&E attendances	0	0	0	0	0	0	0
Social services contacts[d]	0	0	0	0	0	0	0
Antidepressant prescriptions	0.1 (0.0)	0	0	0	−0.0 (−0.1, 0.0)	−0.0 (−0.1, 0.0)	−0.0 (−0.1, 0.0)
Other prescriptions	1.3 (0.1)	1.3 (0.1)	1.3 (0.1)	1.4 (0.1)	0.1 (−0.3, 0.4)	0.0 (−0.3, 0.4)	0.1 (−0.3, 0.5)

Note: Means (cluster-adjusted standard errors and confidence intervals).
[a]Number of baby, mother and PND visits sum to greater than the total number of visits due to some visits being for more than one purpose.
[b]5 or fewer contacts in the control and combined intervention groups. Means for both groups round to zero.
[c]5 or fewer contacts in the control and combined intervention groups; mean rounds to zero.
[d]5 or fewer contacts in the control group and combined intervention groups; mean rounds to zero.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

definition of threshold delineating groups of lower-risk and at-risk women at 6 weeks, considering a cut-off for being at-risk of 10 or more on EPDS and a cut-off of 13 or more (Songoygard *et al.*, 2012; Morrell *et al.*, 2016). We also performed SUR on a two-stage bootstrapped sample of cost and outcomes data (1000 replications) to address potential issues of non-normality in distributions of costs and outcomes (Gomes *et al.*, 2012*a*). Bootstrap sampling was stratified by randomised group. The methods are presented as sensitivity analyses because one caveat to using two-stage bootstrapping is lower-than-nominal coverage probability (Gomes *et al.*, 2012*a*). To examine the impact of missing cases on 6-month results, we created ten complete datasets generated by multilevel multiple-imputation models, and ran them separately for control and intervention groups as previously recommended (Gomes *et al.*, 2013). Models included regressors used in the cost-effectiveness analyses and other baseline factors that predicted missing data on costs and outcomes (feeding method, receipt of benefits, the age of leaving full-time education). Continuous variables were imputed by predictive mean-matching and dichotomous variables were imputed by logistic regression. Multilevel imputation was implemented by chained equations using the *mice* (Buuren and Groothuis-Oudshoorn, 2011) and *miceadds* (Robitzsch *et al.*, 2016) packages in R statistical software (R Core Team, 2016). Results of the SUR from each of the ten complete datasets were combined in Stata using Rubin's rules (Rubin, 1987; StataCorp, 2015).

Results

A total of 2241 lower-risk women (767 control in 37 clusters; 1474 intervention in 63 clusters) completed the EPDS at 6-month follow-up. Data sufficient to compute SF-6D scores at both 6-week and 6-month time-points were available for 2158 participants (736 control in 37 clusters; 1422 intervention in 63 clusters). There were 1459 women with complete economic and SF-6D data at both time-points (417 control in 23 clusters; 1042 intervention in 47 clusters). Participants' baseline characteristics are summarised in Supplementary Table S1.1 for the full lower-risk sample ($N = 2241$), the sample with economic data ($N = 1459$) and the sample without economic data ($N = 782$). Baseline characteristics of the samples with and without economic data within their experimental groups differed in only one respect: at baseline, 3% (10/350) of control group women *without* economic data available reported poor baby health over the previous 4 weeks compared with 1% (3/417) of control group women *with* economic data.

Resource use and costs

Over the 6-month period (Table 2), there were no A&E attendances or admissions to mother and baby psychiatric units, while clinical and community mental health contacts and social services visits were extremely rare (five contacts or fewer in either group). The intervention group had statistically significantly fewer HV visits focused on the mother than the control group (3.6 *v.* 2.0, $p < = 0.001$), with similar results for PND visits (although contacts for this reason, were comparatively low: 0.3 *v.* 0.2). HV contacts that were focused on the mother differed between control and combined (CBA/PCA) intervention sub-groups (3.6 *v.* 1.4, $p = 0.003$). Overall, total HV time spent with the mother/baby was 56 minutes lower in the intervention than control group ($p = 0.027$). Average time spent by HVs in the CBA subgroup was 62 minutes lower ($p = 0.049$) than in the control group. Average time spent by HVs in the PCA subgroup was not different from that in the control group ($p = 0.156$).

Initial bivariate analyses examining a number of visits to women in the 'very low risk' and 'subthreshold risk' groups showed interesting patterns. Mean number of visits related to mother's health, and for PND specifically, appeared to rise as EPDS score rose (online Supplementary Table S2.1). This pattern was not seen in visits related to baby health. A number of visits related to the mother's health was significantly lower in the CBA group compared with control group, and also in the CBA and the PCA groups combined compared to control, for both 'very low risk' and 'subthreshold risk' women. Visits were somewhat but not significantly lower in the PCA group compared with controls within sub-groups.

Further multivariate analyses examined the impact of the PoNDER HV training on the total number of HV visits (for all purposes). In the lower-risk sample, intervention participants received non-significantly more visits than controls (online Supplementary Table S3.1); results were quite similar for CBA and PCA approaches. Analyses also examined whether the intervention had a differential impact on a number of HV visits according to how near mothers' EPDS scores were to 12. Intervention participants in both subthreshold risk and very low-risk groups received (non-significantly) more HV visits than controls, similar to the results over the whole lower-risk sample. The interaction term for combined intervention and risk sub-group was not significantly different from zero ($p = 0.905$); results were similar for CBA and PCA ($p = 0.837$ and $p = 0.647$, respectively). The impact of the intervention appears to have been relatively uniform over the whole of the lower-risk sample.

The overall cost of care for women in the intervention group was significantly lower (difference of £72; 95% CI −137 to −8; cluster-adjusted $t = 2.246$, $p = 0.028$) than for controls (Table 3).

Outcomes

Control and intervention groups were similar in terms of baseline (6-week) utilities (Table 4); utilities were somewhat higher in the intervention group at 6 months. Mean QALYs were significantly higher in the combined intervention group than in controls (mean difference: 0.004 95% CI 0.000–0.008, $p = 0.0466$). There was a 2.8% difference (95% CI −0.5% to 6.1%) in the proportions remaining at low risk at 6 months between the combined CBA and PCA intervention and control groups.

Clustering and correlation of costs and outcomes data

ICCs of QALYs (online Supplementary Table S4.1) were negative and larger in the control group than in the combined intervention group; ICCs for costs were higher in the combined intervention group than in the control group. Mean cluster sizes differed between control (18.1) and intervention groups (22.2). Taking Cohen's criterion (Cohen, 1988) as a gauge of effect size, costs were moderately positively correlated with cluster size in the control group ($r = 0.43$, $p \leqslant 0.001$) and weakly negatively correlated with cluster size in the CBA group ($r = -0.20$, $p \leqslant 0.001$); QALYs were weakly positively correlated with cluster size in the CBA ($r = 0.09$, $p = 0.030$) and weakly negatively correlated with cluster size in the PCA groups ($r = -0.03$, $p = 0.005$); but not correlated when these intervention groups were combined. These

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

Table 3. Costs at 6-months (£), available cases

Item	Control Mean (s.e.) $n = 417$	Intervention (combined) Mean (s.e.) $n = 1042$	Intervention (CBA) Mean (s.e.) $n = 553$	Intervention (PCA) Mean (s.e.) $n = 489$	Intervention (combined) Mean difference from control (95% confidence interval)	Intervention (CBA) Mean difference from control (95% confidence interval)	Intervention (PCA) Mean difference from control (95% confidence interval)
HV contacts	222 (27)	154 (18)	146 (26)	163 (31)	−68 (−132, −4)*	−76 (−156, 4)	−59 (−147, 30)
GP contacts	74 (6)	69 (4)	68 (6)	70 (6)	−5 (−21,10)	−6 (−24, 11)	−4 (−22, 14)
Mother and baby unit admissions	0	0	0	0	0	0	0
Community mental health contacts	0	0	0	0	0	0	0
Clinical mental health contacts	0 (1)	1 (0)	0	1 (1)	1 (−1, 2)	0	1 (−1, 4)
A&E attendances	0	0	0	0	0	0	0
Social services contacts	0	0	0	0	0	0 (−0, 1)	0
Antidepressant prescriptions	0	0	0	0	−0	−0	−0
Other prescriptions	4 (0)	4 (0)	4 (0)	4 (0)	0 (−1, 1)	0 (−1, 1)	0 (−1,1)
Total cost	300 (27)	227 (18)	218 (27)	238 (31)	−72 (−137, −8)*	−82 (−163, −1)*	−61 (−149, 27)

Note: Means (cluster-adjusted standard errors and confidence intervals).
*$p < 0.05$.

results indicate the need to adjust appropriately for both clustering and correlation within the cost-effectiveness analyses.

Results of the cost-effectiveness analyses

The inclusion of covariates with small amounts of missing data slightly decreased the sample available for analysis to 1446 (with the loss of 7/1035 (0.8%) intervention cases and 6/411 (1.4%) controls).

Outcomes

Mean QALY difference between intervention and control groups, adjusted for covariates, was not statistically significant (Table 5). Results are similar for the intervention subgroups, although gains were slightly greater in the CBA than PCA group. For the dichotomised risk of depression outcome, the adjusted difference in the proportions of mothers at low-risk at 6 months was very slightly lower than the unadjusted figures for the combined intervention groups and for CBA and PCA separately.

Costs and cost-effectiveness

Adjusted 6-month costs in the intervention group were £82 lower than in the controls. In the intervention sub-groups, costs in the CBA group were £93 lower and costs in the PCA group £73 lower compared with controls.

The point estimate for the cost of a QALY created by the intervention was negative. A negative ICER may occur when costs are lower and outcomes better in one group (referred to as 'dominance'). In this case, because outcomes were approximately equivalent between groups, although costs were significantly lower in the intervention group, the resulting confidence intervals of the ICER were wide, crossing zero and the upper bound was less than the lower bound (Glick *et al.*, 2007). We can take from the results that the PoNDER intervention is the preferred strategy over usual care, as long as the NHS is willing to pay anywhere from 0 to approximately £66 500 per QALY. The probability of the intervention being cost-effective at £20 000 exceeds 99% (Fig. 1).

When looking at the CEACs for CBA *v.* control and PCA *v.* control, there was little difference between them in the probability of being cost-effective (over 99%) over the range of QALY values between 0 and £20 000 (Fig. 2). CBA had a marginally higher probability of cost-effectiveness, which reflects the slightly lower mean costs of CBA, with similar QALYs gained between all three strategies.

The cost of being in the lower-risk group at 6 months rather than in the at-risk group as a result of the intervention (the ICER) was very low, at about £3500.

Sensitivity analyses

Results of all sensitivity analyses are given in Supplementary Table S5.1.

To examine whether results are robust to violations of the assumption of normally distributed dependent variables, the regressions were applied to data generated by two-stage bootstrapping. There was a 99% probability that the intervention was cost-

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

Table 4. Outcome measures in the sample of cases available for cost-effectiveness analyses

Item	Control Mean (s.e.) *n* = 411	Intervention (combined) Mean (s.e.) *n* = 1035	Intervention (CBA) Mean (s.e.) *n* = 548	Intervention (PCA) Mean (s.e.) *n* = 487	Intervention (combined) Mean difference from control (95% confidence interval)	Intervention (CBA) Mean difference from control (95% confidence interval)	Intervention (PCA) Mean difference from control (95% confidence interval)
6 week utility	0.679 (0.004)	0.685 (0.003)	0.685 (0.004)	0.685 (0.004)	0.006 (−0.004, 0.016)	0.006 (−0.005, 0.017)	0.006 (−0.005, 0.018)
6 month utility	0.831 (0.007)	0.846 (0.004)	0.846 (0.007)	0.846 (0.006)	0.015 (−0.002, 0.031)	0.015 (−0.006, 0.036)	0.015 (−0.003, 0.032)
QALY	0.290 (0.002)	0.294 (0.001)	0.294 (0.002)	0.294 (0.002)	0.004 (0.000, 0.008)*	0.004 (−0.001, 0.009)	0.004 (−0.001, 0.009)
Low-risk status at 6 months[a]	89.1% (1.4%)	91.9% (0.9%)	92.0% (1.3%)	91.8% (1.3%)	2.8% (−0.5%, 6.1%)	2.9% (−01.2, 7.0%)	2.7% (−1.2%, 6.7%)

Note: Means (cluster-adjusted standard errors and confidence intervals).
[a]Binary variable representing whether the participant was at risk of depression, where EPDS⩾12 is coded as 0 and EPDS<12 is coded as 1. The variable is here treated as continuous, and results are expressed in percentage terms.
*$p < 0.05$.

Table 5. Cost-effectiveness analyses: costs, outcomes and ICERs at 6-months

	Group means[a] (s.e.)				Adjusted estimates (Mean difference from control) (95% CI)		
Item	Control *n* = 411	Intervention (combined) *n* = 1035	Intervention (CBA) *n* = 548	Intervention (PCA) *n* = 487	Intervention (combined) *n* = 1446	Intervention (CBA) *n* = 1446	Intervention (PCA) *n* = 1446
QALY[b]	0.292 (0.001)	0.294 (0.001)	0.294 (0.001)	0.294 (0.001)	0.002 (−0.001,0.004)	0.002 (−0.001,0.005)	0.001 (−0.001,0.004)
Costs[c]	311 (23)	229 (12)	218 (12)	238 (19)	−82** (−133, −31)	−93*** (−143, −43)	−73* (−131, −16)
ICER[d]	–	–	–	–	−50 800 (66 500, −14 000)	−52 800 (53 000, −15 500)	−48 900 (50 000, −8 000)
Low-risk status at 6 months[e]	89.4% (1.2%)	91.7% (0.8%)	92.0% (1.2%)	91.4% (0.9%)	2.3% (−1%, 5%)	2.6% (−0.9%, 6%)	2.0% (−1.1%, 5%)
ICER[d]	–	–	–	–	−3 500 (13 800, −1000)	−3 600 (unb'd, unb'd)	−3 700 (unb'd, unb'd)

Note: unb'd = unbounded.
[a]Estimated marginal means, cluster-robust standard errors.
[b]Estimates from SUR equation for QALY adjusted for mother's age, history of PND, living arrangement (alone or with others), any history of major life events, baseline EPDS score, number of other children in the family, whether the mother was economically active, baseline utility.
[c]Estimates from SUR equation for costs adjusted for mother's age, history of PND, living arrangement (alone or with others), any history of major life events, baseline EPDS score, number of other children in the family, whether the mother was economically active, baseline utility.
[d]Rounded to nearest 100.
[e]Binary variable representing whether the participant was at risk of depression, where EPDS⩾12 is coded as 0 and EPDS<12 is coded as 1. The variable is here treated as continuous, and results are expressed in percentage terms. Estimates from SUR equation for low-risk at 6 month adjusted for mother's age, history of PND, living arrangement (alone or with others), any history of major life events, baseline EPDS score, number of other children in the family, whether the mother was economically active.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.01$.

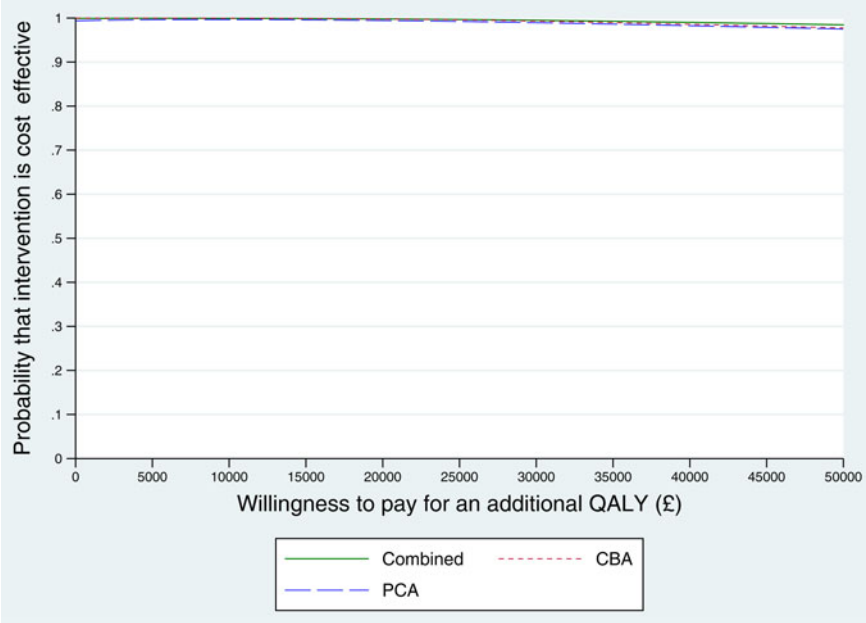


Fig. 1. Cost-effectiveness acceptability curve: QALY.

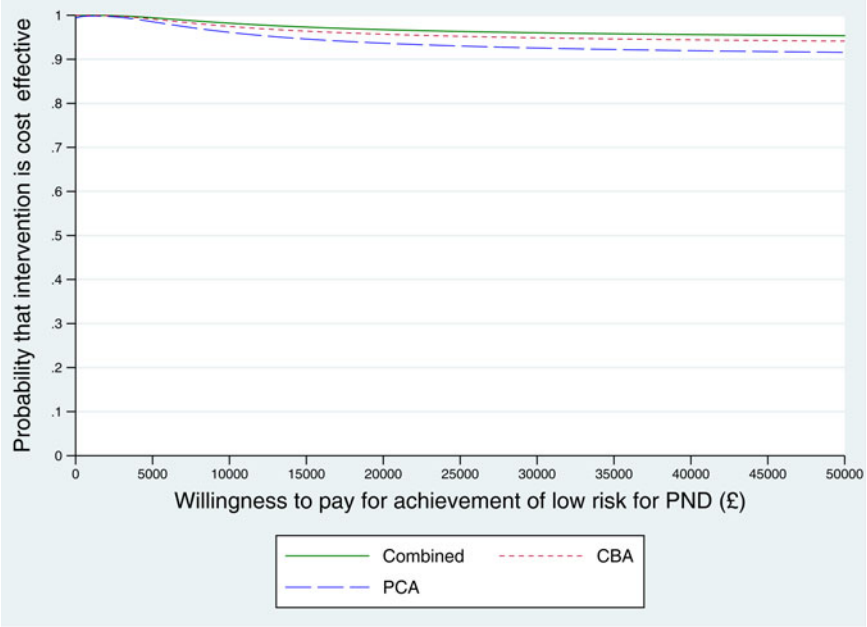


Fig. 2. Cost-effectiveness acceptability curve: dichotomous variable for low-risk of PND at 6 months.

effective at a willingness-to-pay of £20 000 (online Supplementary file S1 Figure S1.1). Results for CBA and PCA were similar, CBA having a 1% higher probability of cost-effectiveness at the £20 000 threshold (99% *v.* 98%).

We varied the cut-off score for the lower-risk sample (considering both a lower cut-off for being at-risk of 10 or more on the EPDS and a higher cut-off of 13 or more on the EPDS) to examine whether the results were robust to changes in the size and composition of the lower-risk group. The estimates of cost and QALY differences were fairly similar to those in the main analyses. With the lower cut-off for higher-risk status, adjusted QALY difference estimates were 0.001 (95% CI −0.002 to 0.004) lower than in the main analyses (0.001 *v.* 0.002). With the higher cut-off, estimates were the same as in the main analyses. Cost differences were again similar to the main analyses for the lower cut-off and £10 lower than in the main analyses with the higher cut-off. CEACs were similar to those in the main analyses (online Supplementary file S1 Figure S1.2).

Running the analyses on imputed data also produced similar results to the main analyses in terms of estimated mean costs and QALYs at 6 months. The adjusted QALY difference was very slightly higher than in the main analysis and the confidence interval did not cross zero; the adjusted cost difference of −£47 (95% CI −85 to −10) was £35 lower than in the main analysis, with a smaller confidence interval. The ICER was considerably reduced (from −50 800 to −16 700), with a confidence interval that was negative, suggesting that, taking sampling uncertainty into account as well as the point estimate, the combined

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

intervention was dominant. CEACs were very similar to those in the main analyses (online Supplementary file S1 Figure S1.3).

Discussion

Our analyses examined the impacts of PoNDER HV training package on 6-month costs and QALYs for mothers at *lower risk* of PND. The intervention for this population of mothers was not only cost-effective at the NICE threshold of £20 000 per QALY gained, but also cost-reducing. This finding is of particular interest given that the eight psychologically-informed HV sessions were primarily targeted at women with an EPDS score of 12 or more (Morrell *et al.*, 2009*b*; Brugha *et al.*, 2011) and not at those 'very low risk' and 'subthreshold risk' women included in the analyses reported here. The choice of approach (CBA or PCA) made relatively little difference to cost or to the probability of cost-effectiveness at the £20 000 threshold, suggesting that training in CBA and PCA approaches had more or less equivalent economic consequences. The impact of PoNDER HV training did not appear to be confined to those women closer to the EPDS threshold score of 12, as evidenced by analyses of a number of HV visits across risk sub-groups and by sensitivity analyses varying the threshold for lower-risk.

In relation to the secondary outcome, the 2.3% mean difference between intervention and control groups in proportions of lower-risk women at 6 months was not statistically significant. In analyses elsewhere of a larger sample of lower-risk women ($N = 2241$) than the 1446 observations available for the economic analysis, the percentage of women with an EPDS score of 12 or more 6-months postnatally was 10.8% of control women and 7.7% of intervention women, a difference of 3.1% (95% CI 0.4–5.9%) (Morrell *et al.*, 2009*a*; Brugha *et al.*, 2011). Nonetheless, in the economic evaluation sample, the probability of cost-effectiveness was very high over a range of willingness-to-pay thresholds below £14 000 for being lower-risk rather than at-risk at 6 months.

These findings provide strong evidence that the training programme was cost-effective in preventing depression 6 months after childbirth in mothers at *lower risk* of depression even though psychological intervention sessions were not targeted on them (Brugha *et al.*, 2011). This intervention reduced the risk of depression and paid for itself over 6 months.

We also found that the number of visits related to the mother's health was significantly lower in the combined intervention group compared with controls. After the training, the intervention groups HVs were more confident in assessing risk and reassessing women than the control group HVs (Morrell *et al.*, 2009*b*). The HVs offered face-to-face psychological support sessions to women who were indicated as depressed according to their clinical assessment and face-to-face EPDS score. The HVs could distinguish true depression from extreme tiredness and labile mood. Therefore, the intervention group HVs appropriately responded to all levels of risk according to the combination of the EPDS score and enhanced clinical assessment skills gained during the training. This may have made them more efficient in their visits to those women at greater risk rather than to those at less risk in comparison with the control groups. In contrast, control group HVs were not trained to have the skills to know which women were truly depressed and therefore may have visited those women who had extreme tiredness and some symptoms of depression but were not depressed.

Strengths and limitations

Strengths of the study include a large number of observations available for analysis, analytic methods appropriate to clustered data and sensitivity analyses of key assumptions in those analyses. The purpose of cluster allocation in training interventions is to protect against contamination of the untrained control group; disadvantages of this approach include potential selection biases during recruitment, increased complexity of design and increased sample sizes required compared with individual randomisation (Klar, 2015). Cluster allocation also ensured that health outcomes related to all HVs within each cluster, thereby strengthening the generalisability of results. Our analyses addressed imbalances between trial arms in terms of numbers and size of clusters, an issue not considered in the original analyses (Morrell *et al.*, 2009*b*). Other potential challenges to the robustness of results (impacts of missing cost data and skewness of dependent variables) were explored in sensitivity analyses; they did not make any difference to the conclusions drawn from the main analyses.

There were several limitations to this study. Data available for the cost-effectiveness analysis were less complete than those available on health outcome data; however, sensitivity analyses drawing on multiply-imputed data confirmed the findings, indicating if anything even stronger evidence of the dominance of the intervention over usual care.

Cost measures were confined to health and social care services, a limitation when looking at women at lower risk of depression who have less need for support. We only analysed costs over 6-months postnatally so we did not factor in the risk of longer-term adverse effects on child development and associated costs (Bauer *et al.*, 2016*b*) or on employment-related productivity losses associated with depression (Thomas and Morris, 2003).

Present-day unit costs may differ from those used here (2003/04 prices) which reflect the organisation of care and mix of services at the time of the trial. Organisation of care, outcomes and unit costs were undoubtedly interrelated and cannot now be easily disentangled. The organisation of health visiting has changed over time: HVs are registered nurses or midwives who can now gain additional qualifications to become specialist community public health nurses (Cowley *et al.*, 2013; Health Education England, 2016). There has been a national programme to improve access to child health services, increasing numbers of HVs, and to transfer the commissioning of health visiting to local government (Department of Health, 2015). Such changes could affect unit costs. The cost of an hour of health visiting (client-related work) in 2014/15 was £76 (Curtis and Burns, 2015), whereas the hourly cost used here would be £101 if uprated to 2014/15 prices (Curtis and Burns, 2015). Data sources and methods used to estimate HV unit costs have changed since 2004 (Netten and Curtis, 2004), as no data on HV time-use was available for later calculations (e.g. for face-to-face/indirect contacts and travel) (Curtis and Burns, 2015).

Our cost-effectiveness analysis adopted methods consistent with good practice guidelines (Ramsey *et al.*, 2015) and employed methods relevant to clustered data, but the choice of analytical model can influence results (Mantopoulos *et al.*, 2016) and so is a source of methodological uncertainty. However, given the strength of our conclusions, such uncertainties are unlikely to be a concern.

In the base-case analysis, the mean utility scores of intervention group mothers were not significantly greater than for control group mothers. We might ask whether measurement of changes in utility in mothers at lower risk for depression can be as accurate

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

as in an at-risk population. However, the SF-6D is sensitive to differences in EPDS scores, discriminating well between different dichotomised levels of risk for PND (Petrou *et al.*, 2009).

Implications for policy and practice

PoNDER HV training offers the benefit of a service delivered in routine postnatal care with an assessment by HVs with whom women are already in contact: it is a low-cost universal preventive intervention. It reduces the risk of developing PND symptoms, reduces health and social care service use over 6-months and is cost-effective. Recent reviews on the cost-effectiveness of preventive and early interventions (Bauer *et al.*, 2016*a*; Morrell *et al.*, 2016) suggest that to date there are no other reports of an economic evaluation alongside a clinical trial to *prevent* perinatal mental health problems. Decision models drawing on the economic evidence have found that some interventions that address mild or subthreshold symptoms (including PCA and CBT-based universal approaches) are likely to be cost-effective and in some cases also lead to cost savings (Morrell *et al.*, 2016). There is also evidence that assessment by trained professionals such as HVs can lead to better outcomes for postnatal women including reduced risk of depression (Bauer *et al.*, 2016*b*; O'Connor *et al.*, 2016).

Two major global challenges in relation to mental illness are the 'treatment gap'(Kohn *et al.*, 2004) and 'prevention gap' (Jorm *et al.*, 2017). Rates of undiagnosed and untreated PND are particularly high (Bijl *et al.*, 2003), yet many women with perinatal depression do not take up screening (Reay *et al.*, 2011). One implication of our study is that there need not be a 'prevention gap': women at lower risk of depression would benefit from support by HVs additionally trained in assessment and psychological support. A universal prevention programme of this kind would come at no extra cost to the healthcare system; indeed it would be cost-reducing. There are potential implications for women's perception of available support should they need it (Henderson *et al.*, 1981; Brugha *et al.*, 1998): women with perinatal depression can be fearful of accessing mental health services (Slade *et al.*, 2010), for instance worrying that their children will be taken into care (Dolman *et al.*, 2013; Megnin-Viggars *et al.*, 2015).

Conclusion

Our analyses confirm that PoNDER HV training in assessment for symptoms of PND plus the skills to provide a psychologically informed intervention (CBA or PCA) is cost-effective, even when additional psychological care is not indicated. This provides support for further investigation of the merits of a universal service that includes extra HV training in clinical assessment and the ability to offer psychological support if indicated.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291718001940

Clinical trial registration number. ISRCTN92195776 (www.controlled-trials.com/ISRCTN92195776)

Financial support. The PoNDER trial was funded by NHS Health Technology Assessment, England; CH and MK inputs to the design and economic analysis reported in this paper were funded by an NIHR Senior Investigator award to MK.

Conflict of interest. MK and CH reported grants from NIHR during the conduct of the study. The authors declare no conflicts of interest.

Ethics committee approval. Trent multicentre research ethics committee.

References

Adams G, Gilliford M, Ukoumunne O, Eldridge S, Chinn S and Campbell M (2004) Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *Journal of Clinical Epidemiology* **57**, 785–794.

Appleby L, Warner R, Whitton A and Faragher B (1997) A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. *British Medical Journal Publishing Group BMJ (Clinical research ed.)* **314**, 932–936.

Bauer A, Knapp M and Adelaja B (2016*a*) *Best Practice for Perinatal Mental Health Care: The Economic Case*. London: Personal Social Services Research Unit.

Bauer A, Knapp M and Parsonage M (2016*b*) Lifetime costs of perinatal anxiety and depression. *Journal of Affective Disorders* **192**, 83–90.

Bijl RV, De Graaf R, Hiripi E, Kessler RC, Kohn R, Offord DR, Ustun TB, Vicente B, Vollebergh WAM, Walters EE and Wittchen HU (2003) The prevalence of treated and untreated mental disorders in five countries. *Health Affairs* **22**, 122–133.

Brazier J, Roberts J and Deverill M (2002) The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics* **21**, 271–292.

British Medical Association and Royal Pharmaceutical Society (2011) *British National Formulary* (BNF). London: BMA and RPS.

Brugha TS, Morrell CJ, Slade P and Walters SJ (2011) Universal prevention of depression in women postnatally: cluster randomized trial evidence in primary care. *Psychological Medicine* **41**, 739–748.

Brugha TS, Sharp HM, Cooper S-A, Weisender C, Britto D, Shinkwin R, Sherrif T and Kirwan PH (1998) The Leicester 500 Project. Social support and the development of postnatal depressive symptoms, a prospective cohort survey. *Psychological Medicine* **28**, 63–79.

Buuren S van and Groothuis-Oudshoorn K (2011) Mice: multivariate imputation by chained equations in *R*. *Journal of Statistical Software* **45**, 1–67.

Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cowley S, Caan W, Dowling S and Weir H (2007) What do health visitors do? A national survey of activities and service organisation. *Public Health* **121**, 869–879.

Cowley S, Whittaker K, Grigulis A, Malone M, Donetto S, Wood H, Morrow E and Maben J (2013) *Why Health Visiting? A Review of the Literature About key Health Visitor Interventions, Processes and Outcomes for Children and Families*. London: National Nursing Research Unit, King's College London. National Nursing Research Unit, King's College London.

Cox JL, Holden JM and Sagovsky R (1987) Detection of postnatal depression: development of the 10-item Edinburgh postnatal depression scale. *British Journal of Psychiatry* **150**, 782–786.

Curtis L and Burns A (2015) *Unit Costs of Health and Social Care 2015*. Canterbury: Personal Social Services Research Unit (PSSRU): University of Kent.

Davis WW (2001) *Design and Analysis of Cluster Randomization Trials in Health Research*. Journal of the American Statistical Association vol 96. London: Arnold.

Department of Health (2000) *NHS Trust Financial Returns*. Leeds: Department of Health.

Department of Health (2005) *NHS reference Costs 2004*. Leeds: Department of Health.

Department of Health (2015) Transfer of 0–5 children's public health commissioning to Local Authorities.

Dolman C, Jones I and Howard LM (2013) Pre-conception to parenting: a systematic review and meta-synthesis of the qualitative literature on motherhood for women with severe mental illness. *Archives of Women's Mental Health* **16**, 173–196.

Ford E, Shakespeare J, Elias F and Ayers S (2017) Recognition and management of perinatal depression and anxiety by general practitioners: a systematic review. *Family Practice* **34**, 11–19.

Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G and Swinson T (2005) Perinatal depression. *Obstetrics & Gynecology* **106**, 1071–1083.

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940

Gavin NI, Meltzer-Brody S, Glover V and Gaynes BN (2015) Is population-based identification of perinatal depression and anxiety desirable? In Milgrom J and Gemmill AW (eds), *Identifying Perinatal Depression and Anxiety: Evidence-Based Practice in Screening, Psychosocial Assessment, and Management*. Chichester, UK: John Wiley & Sons, Ltd, pp. 11–31.

Glick HA, Doshi JA, Sonnad SS and Polsky D (2007) *Economic Evaluation in Clinical Trials*. *OUP Catalogue*. Oxford, New York: Oxford University Press.

Gomes M, Díaz-Ordaz K, Grieve R and Kenward MG (2013) Multiple imputation methods for handling missing data in cost-effectiveness analyses that use data from hierarchical studies: an application to cluster randomized trials. *Medical Decision Making* **33**, 1051–1063.

Gomes M, Grieve R, Nixon R and Edmunds WJ (2012*a*). Statistical methods for cost-effectiveness analyses that use data from cluster randomized trials: a systematic review and checklist for critical appraisal. *Medical Decision Making* **32**, 209–220.

Gomes M, Grieve R, Nixon R, Ng ESW, Carpenter J and Thompson SG (2012*b*). Methods for covariate adjustment in cost-effectiveness analysis that use cluster randomised trials. *Health Economics* (*United Kingdom*) **21**, 1101–1118.

Hansard (2004*a*). Written answers, Monday 23 May 2005.

Hansard (2004*b*). Written answers, Tuesday 19 October 2004.

Health Education England (2016) *Explore roles: Health visitor*.

Henderson S, Byrne DG and Duncan-Jones P (1981) *Neurosis and the Social Environment*. Sydney, London: Academic Press.

Herrin J (2012) CLTEST: Stata modules for performing cluster-adjusted chi-square and *t* tests. Statistical Software Components.

Hewitt CE and Gilbody SM (2009) Is it clinically and cost effective to screen for postnatal depression: a systematic review of controlled clinical trials and economic evidence. *BJOG: An International Journal of Obstetrics and Gynaecology* **116**, 1019–1027.

Holden JM, Sagovsky R and Cox JL (1989) Counselling in a general practice setting: controlled study of health visitor intervention in treatment of postnatal depression. *British Medical Journal Publishing Group BMJ (Clinical research ed.)* **298**, 223–226.

Jorm AF, Patten SB, Brugha TS and Mojtabai R (2017) Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry* **16**, 90–99.

Klar N (2015) *A Practical Guide to Cluster Randomised Trials in Health Services Research*, vol. **3406**. Chichester: Wiley.

Kohn R, Saxena S, Levav I and Saraceno B (2004) The treatment gap in mental health care. *Bulletin of the World Health Organization* **82**, 858–866.

Manca A, Hawkins N and Sculpher MJ (2005*a*). Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Economics* **14**, 487–496.

Manca A, Rice N, Sculpher MJ and Briggs AH (2005*b*). Assessing generlisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Economics* **14**, 471–485.

Mantopoulos T, Mitchell PM, Welton NJ, McManus R and Andronis L (2016) Choice of statistical model for cost-effectiveness analysis and covariate adjustment: empirical application of prominent models and assessment of their results. *European Journal of Health Economics* **17**, 927–938.

Megnin-Viggars O, Symington I, Howard LM and Pilling S (2015) Experience of care for mental health problems in the antenatal or postnatal period for women in the UK: a systematic review and meta-synthesis of qualitative research. *Archives of Women's Mental Health* **18**, 745–759.

Milgrom J and Gemmill AW (2014) Screening for perinatal depression. *Best Practice and Research Clinical Obstetrics and Gynaecology* **28**, 13–23.

Morrell CJ, Ricketts T, Tudor K, Williams C, Curran J and Barkham M (2011) Training health visitors in cognitive behavioural and person-centred approaches for depression in postnatal women as part of a cluster randomised trial and economic evaluation in primary care: the PoNDER trial. *Primary Health Care Research & Development* **12**, 11–20.

Morrell CJ, Slade P, Warner R, Paley G, Dixon S, Walters SJ, Brugha T, Barkham M, Parry GJ and Nicholl J (2009*a*). Clinical effectiveness of health visitor training in psychologically informed approaches for depression in postnatal women: pragmatic cluster randomised trial in primary care. *BMJ* **338**, a3045.

Morrell CJ, Sutcliffe P, Booth A, Stevens J, Scope A, Stevenson M, Harvey R, Bessey A, Cantrell A, Dennis CL, Ren S, Ragonesi M, Barkham M, Churchill D, Henshaw C, Newstead J, Slade P, Spiby H and Stewart-Brown S (2016) A systematic review, evidence synthesis and meta-analysis of quantitative and qualitative studies evaluating the clinical effectiveness, the cost-effectiveness, safety and acceptability of interventions to prevent postnatal depression. *Health Technology Assessment* **20**, 1–414.

Morrell CJ, Warner R, Slade P, Dixon S, Walters S, Paley G and Brugha T (2009*b*). Psychological interventions for postnatal depression: cluster randomised trial and economic evaluation. The PoNDER trial. *Health Technology Assessment* **13**, iii–iv, xi–xiii, 1–153.

National Institute for Clinical Excellence (2004) *Guide to The Methods of Technology Appraisal*. London: National Institute for Clinical Excellence.

National Institute for Health and Care Excellence (2013) *Guide to The Methods of Technology Appraisal 2013*. London: National Institute for Health and Care Excellence.

National Institute for Health and Clinical Excellence (2008) *Guide to The Methods of Technology Appraisal*. London: National Institute for Health and Clinical Excellence.

Netten A and Curtis L (2004) *Unit Costs of Health and Social Care 2004*. Canterbury: Personal Social Services Research Unit (PSSRU): University of Kent.

O'Connor E, Rossom RC, Henninger M, Groom HC and Burda BU (2016) Primary care screening for and treatment of depression in pregnant and postpartum women. *American Medical Association* JAMA **315**, 388.

Panageas KS, Schrag D, Russell Localio A, Venkatraman ES and Begg CB (2007) Properties of analysis methods that account for clustering in volume–outcome studies when the primary predictor is cluster size. *Statistics in Medicine* **26**, 2017–2035.

Petrou S, Morrell J and Spiby H (2009) Assessing the empirical validity of alternative multi-attribute utility measures in the maternity context. *Health and Quality of Life Outcomes* **7**, 40.

Prady SL, Pickett KE, Petherick ES, Gilbody S, Croudace T, Mason D, Sheldon TA and Wright J (2016) Evaluation of ethnic disparities in detection of depression and anxiety in primary care during the maternal period: combined analysis of routine and cohort data. *British Journal of Psychiatry* **208**, 453–461.

Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, Briggs A and Sullivan SD (2015) Cost-effectiveness analysis alongside clinical trials II—an ISPOR good research practices task force report. *Value in Health* **18**, 161–172.

R Core Team (2016) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reay R, Matthey S, Ellwood D and Scott M (2011) Long-term outcomes of participants in a perinatal depression early detection program. *Journal of Affective Disorders* **129**, 94–103.

Robitzsch A, Grund S and Henke T (2016) Some additional multiple imputation functions, especially for mice (Version 2.1-0).

Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys*. New York, Chichester: Wiley.

Slade P, Morrell CJ, Rigby A, Slade P, Morrell CJ, Rigby A and Ricci K (2010) Postnatal women' s experiences of management of depressive symptoms: a qualitative study. *The British Journal of General Practice: the Journal of the Royal College of General Practitioners* **60**, 440–448.

Songoygard KM, Stafne SN, Evensen KA, Salvesen KA, Vik T and Morkved S (2012) Does exercise during pregnancy prevent postnatal depression? A randomized controlled trial. *Acta Obstetricia et Gynecologica Scandinavica* **91**, 62–67.

StataCorp (2015) *Stata Statistical Software: Release 14 SE*. College Station, Texas: StataCorp LP.

Thomas CM and Morris S (2003) Cost of depression among adults in England in 2000. *British Journal of Psychiatry* **183**, 514–519.

Ukoumunne OC (2002) A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. *Statistics in Medicine* **21**, 3757–3774.

Downloaded from https://www.cambridge.org/core. IP address: 88.212.164.88, on 30 Nov 2018 at 09:51:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0033291718001940