Published in final edited form as:

Clin Psychol Rev. 2011 July ; 31(5): 839–849. doi:10.1016/j.cpr.2011.03.009.

A meta-analysis of treatments for perinatal depression

Laura E. Sockol^{a,*}, C. Neill Epperson^{b,c}, and Jacques P. Barber^d

^aDepartment of Psychology, University of Pennsylvania, United States

^bDepartment of Psychiatry, University of Pennsylvania, United States

^cDepartment of Obstetrics/Gynecology, University of Pennsylvania, United States

^dDepartment of Psychiatry, University of Pennsylvania and Philadelphia VA Medical Center, United States

Abstract

This meta-analysis assessed efficacy of pharmacologic and psychological interventions for treatment of perinatal depression. A systematic review identified 27 studies, including open trials ($n=9$), quasi-randomized trials ($n=2$), and randomized controlled trials ($n=16$) assessing change from pretreatment to posttreatment or comparing these interventions to a control group. Uncontrolled and controlled effect sizes were assessed in separate meta-analyses. There was significant improvement in depressive symptoms from pretreatment to posttreatment, with an uncontrolled overall effect size (Hedges' g) of 1.61 after removal of outliers and correction for publication bias. Symptom levels at posttreatment were below cutoff levels indicative of clinically significant symptoms. At posttreatment, intervention groups demonstrated significantly greater reductions in depressive symptoms compared to control groups, with an overall controlled effect size (Hedges' g) of 0.65 after removal of outliers. Individual psychotherapy was superior to group psychotherapy with regard to changes in symptoms from pretreatment to posttreatment. Interventions including an interpersonal therapy component were found to have greater effect sizes, compared to control conditions, than interventions including a cognitive-behavioral component. Implications of the findings for clinical practice and future research are discussed.

Keywords

Postpartum; Psychotherapy; Pregnancy; Treatment; Depression; Meta-analysis

Perinatal depression is one of the most common complications of childbearing. Approximately 10 to 15% of women experience a clinically significant major depressive episode during pregnancy or the early postpartum period (Bennett, Einarson, Taddio, Koren, & Einarson, 2004b; Epperson, 1999; Gavin, Gayner, Lohr, Meltzer-Brody, Gartlehner and Swinson, 2005; O'Hara & Swain, 1996). These prevalence estimates predominantly reflect rates of depressive symptoms in developed countries; there is evidence that rates of

depression vary more widely in non-developed countries (Halbreich & Karkun, 2006). In addition to the distress and impairment experienced by depressed women, depression during this time period is associated with further adverse outcomes for both mother and child. Women who experience perinatal depressive episodes are at increased risk for subsequent episodes of both postpartum and non-postpartum depression (Cooper & Murray, 1995). Prenatal depression is associated with increased risk for negative birth outcomes, including preterm labor, low birthweight, and intrauterine growth restriction (Grote, Bridge, Gavin, Melville, Iyengar and Katon, 2010). Maternal depression during the postpartum period is also a risk factor for a range of adverse child outcomes, including behavioral problems and impaired cognitive development (Grace, Evindar, & Stewart, 2003).

Given the prevalence of perinatal depression and the adverse effects this disorder has on women and their children, the identification of effective treatments for this disorder has important public health implications. Although there is a great deal of evidence for the efficacy of both antidepressant medication and psychological interventions for depression (see Joffe, Sokolov, & Streiner, 1996 and Cuijpers, van Straten, Andersson, & van Oppen, 2008 for meta-analyses of the efficacy of antidepressant medication and psychotherapy for depression, respectively), concerns unique to the perinatal period may influence the efficacy of these treatments for this population (Kim, O'Reardon, & Epperson, 2010). For example, in an attempt to limit fetal exposure, antidepressants may be prescribed below therapeutic dosage levels (Bennett, Einarson, Taddio, Koren, & Einarson, 2004a). This problem is complicated further by the fact that most women actually require higher doses of antidepressant medication during pregnancy (Dawes & Chowieńczyk, 2001; Hostetter, Stowe, Strader, McLaughlin, & Llewellyn, 2000; Wisner, Perel, & Wheeler, 1993). Concerns regarding effects of infant exposure to antidepressant medication via breast milk may also lead clinicians to prescribe inadequate doses of these medications during the postpartum period (Epperson, Anderson, & McDougale, 1997; Epperson, Jatlow, Czarkowski, & Anderson, 2003). Biological and psychosocial changes that occur in the context of pregnancy and parenting, including sleep deprivation, disruptions to the hormonal milieu, alterations to HPA axis functioning, and changes to interpersonal relationships, introduce challenges that may affect the efficacy of both pharmacological and psychotherapeutic interventions (Dennis & Ross, 2005; Kammerer, Taylor, & Glover, 2006). Efficacy of psychological interventions for depression may also be reduced among women who have had previous pregnancy losses, complications, or traumatic deliveries, as they may experience post-traumatic stress disorder or other comorbid anxiety disorders (Forray, Mayes, Magriples, & Epperson, 2009). There are also differences in the acceptability of interventions, particularly among women who are pregnant or breastfeeding: the majority of women indicate a preference for psychological interventions to antidepressant medication during both pregnancy and the postpartum period, and the overall acceptability of antidepressant medication among these groups is low (Chabrol, Teissedre, Armitage, Danel, & Walburg, 2004; Kim et al., 2011). Thus identification of efficacious interventions, particularly psychological interventions, for this population is an important and growing area of research.

Two meta-analytic reviews of psychological treatments for postpartum depression have found these interventions to be superior to routine care or control conditions. A Cochrane

review of psychological and psychosocial interventions for postpartum depression found that, compared to routine care, these interventions were associated with a 30% reduction in relative risk for depressive symptomatology (Dennis & Hodnett, 2007). Cuijpers, Brännmark, and van Straten (2008) conducted a meta-analysis of 17 studies in which a psychological intervention initiated during the postpartum period was compared to a control or active treatment condition. They reported that psychological interventions were superior to control conditions, with an overall effect size in the moderate range.

In the only published meta-analysis to include interventions for both antenatal and postpartum depression, Bledsoe and Grote (2006) evaluated efficacy of 16 psychological and pharmacological interventions and found that depressive symptoms decreased significantly from pre- to posttreatment. They did not assess the effect of treatments compared to control conditions. While their findings provide preliminary evidence for efficacy of these interventions, a major limitation of their meta-analysis is the difficulty of interpreting effect sizes representing changes in symptoms from pretreatment to posttreatment. As there is evidence that, for many women, depressive symptoms remit naturally over the course of the postpartum period, it is not possible to determine whether these effect sizes reflect the effects of the interventions or simply natural decreases in symptom levels over time (Heron, O'Connor, Evans, Golding, & Glover, 2004). The authors also did not report analyses of homogeneity, tests for outliers or publication bias, did not specify whether the analysis was conducted using fixed or random effects models, and included multiple effect sizes for three studies in which more than one active intervention was assessed. As these methodological issues may substantially impact estimation of effect sizes, these results should be interpreted with caution.

The present meta-analysis addresses several of the limitations of the above studies. Unlike Cuijpers, Brännmark, and van Straten (2008), we included pharmacological interventions in addition to psychological interventions, and included interventions initiated during pregnancy. As many studies of interventions for this population are either open trials or do not include a no-treatment control condition, we elected not to restrict these analyses to studies in which interventions were compared to a control condition. However, to address the possibility that effect sizes calculated from these studies may reflect natural symptom remission over time, we also compared active treatments to control conditions in studies where it was possible to do so. We have also included several new studies of treatments for perinatal depression that have been published since these earlier meta-analyses were conducted. The goal of the current meta-analysis was to assess the efficacy of psychological and pharmacological interventions for perinatal depression, defined as the period encompassing pregnancy and the first 12 months postpartum. Both the overall effect of these interventions on depressive symptoms over time and the relative efficacy of interventions compared to control conditions were assessed. We also conducted exploratory moderator analyses assessing elements of both study design and interventions as potential moderators of the magnitude of effect size when significant heterogeneity of effect sizes was observed.

1. Method

1.1. Search procedures and selection of studies

Relevant studies were identified through searches of databases through September 2010, including PubMed and PsycInfo, using the following terms as descriptors: *postpartum depression, pregnancy and depression, therapy, drug therapy, cognitive behavior therapy, interpersonal psychotherapy, psychodynamic therapy, treatment, and treatment outcome*. The reference lists of existing meta-analyses, reviews, chapters, and retrieved articles were inspected for further studies. Clinical trial databases (including the Cochrane Pregnancy and Childbirth Group, Cochrane Depression, Anxiety and Neurosis Group, and the International Standard Randomized Controlled Trial Number Register) were also reviewed for eligible studies.

To be included in the meta-analysis, studies had to meet the following inclusion criteria:

- a. Used a prospective pretreatment–posttreatment, quasi-randomized trial or randomized controlled trial design.
- b. Assessed impact of antidepressant treatment or specified/manualized psychological intervention for perinatal depression. Hormonal pharmacological interventions, such as estrogen therapy, were excluded. Nonspecific psychosocial interventions, such as peer support groups, were excluded. Interventions that did not explicitly target depressive symptoms, such as smoking cessation programs, were also excluded.
- c. Subjects were limited to women with unipolar depression (defined by diagnostic criteria or symptom severity) during pregnancy or the postpartum period (defined as the 12 months following the birth of a child).
- d. Reported outcomes for depressive symptoms using a validated self-report or clinician-administered measure.
- e. Reported sufficient outcomes to allow for calculation of either uncontrolled or controlled effect sizes.

A flow chart depicting the search process and exclusion of studies is presented in Fig. 1. After removal of duplicates, the search procedure yielded 1447 studies. The 152 studies whose abstracts indicated evaluation of an intervention for antenatal or postnatal depression were obtained and reviewed for inclusion. Of these 152 studies, 122 were excluded for the following reasons: described an intervention without reporting results of an evaluation ($n=10$), study design was not a prospective pretest–posttest, quasi-randomized or randomized controlled trial (e.g., retrospective chart reviews, $n=14$), prevention studies that included women without elevated depressive symptoms or a diagnosis of depression ($n=30$), no pharmacological or psychological intervention (e.g., exercise, hormonal, and social support interventions, $n=47$), population was not restricted to women during the perinatal period or with unipolar depression ($n=7$), studies that reported only qualitative data ($n=3$) or insufficient data for the calculation of effect sizes ($n=8$), and secondary sources for included studies that did not report outcomes relevant to the analyses ($n=3$). The remaining 30 articles, representing 27 studies, were eligible for inclusion in the meta-analysis. When a

secondary source was available for a given study, the secondary source (e.g., Cooper & Murray, 1997; Goodman, Broth, Hall, & Stowe, 2008; Murray, Cooper, Wilson, & Romaniuk, 2003) was used to calculate the effect size when reported data in the primary source was insufficient. Sufficient outcome measures for calculation of effect sizes representing change from pretreatment to posttreatment were reported in 25 studies, and 14 studies reported sufficient outcome measures for calculation of effect sizes representing the difference between treatment and control conditions at posttreatment.

1.2. Coding of studies

All studies were coded for: intervention type (antidepressant medication vs. psychotherapy vs. combined), study design (open trial vs. quasi-randomized trial vs. randomized controlled trial), type of control group (treatment as usual vs. enhanced treatment as usual vs. waiting list vs. active), population (antepartum vs. postpartum), outcome measure, whether the study required a clinician-verified diagnosis of depression for inclusion, treatment length (weeks), and percent attrition. Studies including a psychological intervention were also coded for therapeutic orientation, whether therapy was conducted individually or in a group format, and the location in which therapy was administered (clinic vs. home vs. school). As a majority of studies included the Edinburgh Postnatal Depression Scale (EPDS) as an outcome measure, this measure was used to calculate effect sizes for all studies reporting EPDS outcomes. For studies that did not include the EPDS as an outcome measure, or for which effect sizes could not be calculated using the reported EPDS values, the primary outcome measure was used.

Effect sizes and moderators were coded by the first author. Three of the variables included in the moderator analyses were also coded by a second rater, who was trained in the coding scheme and utilized a written coding manual. Observed agreement was 27/27 for each variable, or 100%. As all of the variables that were coded for use in the moderator analyses were objective variables that were explicitly specified in the studies and required minimal judgment on the part of the coder, this reliability check was considered adequate.

1.3. Analyses

Two separate analyses were conducted. The first analysis assessed the change in depressive symptoms from pretreatment to posttreatment using the standardized mean gain score for all treatment groups. To differentiate these analyses from those comparing treatment to control conditions, these within group effect sizes will be referred to as “uncontrolled effect sizes” (Feske & Chambless, 1995). To ensure independence of included effect sizes, a single effect size was calculated on the basis of the overall mean and standard deviation of the total group of treated subjects in studies that included more than one active treatment. Uncontrolled effect sizes were calculated by dividing the mean change from pretreatment to posttreatment by the pooled standard deviation of the difference score, corrected for upward bias using Hedges' g (Hedges, 1981):

$$g = c_m \left[\frac{M_{Post} - M_{Pre}}{SD_{Diff} / \sqrt{(2(1 - r))}} \right]$$

where the pooled standard deviation is defined as

$$SD_{Diff} = \sqrt{SD_{Pre}^2 + SD_{Post}^2 - 2rSD_{Pre}SD_{Post}}$$

and

$$c_m = 1 - \frac{3}{4(n-1)}.$$

Uncontrolled effect sizes were calculated so that positive effect sizes represented a decrease in depressive symptoms from pretreatment to posttreatment.

None of the studies included in the meta-analysis reported the pretest–posttest correlation for the sample or data that would allow this value to be calculated. Following the recommendations of Lipsey and Wilson (2001), the test–retest reliability of the measures was used as a proxy for the pretest–posttest correlation. These values were estimated from published validation studies of each measure; when multiple test–retest reliabilities were available for a single measure, r was computed as the weighted mean of the reliabilities. As it is likely that these values are inflated estimates of the pretest–posttest correlation, the overall analyses were also conducted using values of 0.3, 0.5, and 0.8 as estimates of low, medium, and high correlations, respectively. There were no substantive differences between these effect sizes, suggesting that using the test–retest correlations as a proxy for this value would not impact the results of the analyses.

The second analysis compared efficacy of active treatments to control conditions using the standardized mean group difference. To differentiate these effect sizes from those defined previously, these between group effect sizes will be referred to as “controlled effect sizes” (Feske & Chambless, 1995). Means and standard deviations for the total group of treated subjects were calculated for all studies in which multiple active treatments were compared to a control group in order to ensure the independence of effect sizes. Controlled effect sizes were calculated by dividing the difference between treatment and control means by the pooled standard deviation, corrected for upward bias using Hedges' g (Hedges, 1981):

$$g = c_m \left[\frac{M_T - M_C}{SD_P} \right]$$

where the pooled standard deviation is defined as

$$SD_P = \sqrt{\frac{(n_T - 1)SD_T^2 + (n_C - 1)SD_C^2}{(n_T + n_C - 2)}}$$

and c_m is defined as described above. Controlled effect sizes were calculated so that positive effects represented lower scores in the intervention group compared to the control group.

Heterogeneity of effect sizes was examined using the Q statistic and the I^2 index. Significant Q statistics indicate that the observed range of effect sizes is significantly larger than would be expected based on within-study variance. While a significant Q statistic indicates heterogeneous effect sizes, nonsignificant Q statistics should be interpreted with caution, as heterogeneous effect sizes may yield a nonsignificant Q value due to low power. The I^2 value indicates the proportion of variance in effect sizes accounted for by between-study variance. The index has a range from 0 to 100; Higgins, Thompson, Deeks, and Altman (2003) suggest that 25, 50 and 75% I^2 values indicate low, medium and high levels of heterogeneity, respectively.

When analyses indicated significant heterogeneity among effect sizes, exploratory analyses were conducted to assess for moderators of effect size. Categorical moderators were assessed using an analysis of variance (ANOVA) of mixed-effects models for each variable hypothesized to influence the effect size. Meta-regression analyses were conducted to assess the effects of continuous moderators. Two types of moderators were included in the analyses. The first were variables that reflected elements of the research design of included studies; significant findings of moderation would indicate that differences in effect sizes could be attributed to methodological variability among studies (e.g., whether studies required a clinician-verified diagnosis for inclusion). The second were variables related to characteristics of interventions; for example, pharmacotherapy vs. psychotherapy and differences between psychological interventions of different therapeutic orientations. For studies assessing psychotherapeutic interventions, we also assessed whether characteristics of the intervention (including the mode of administration, location the intervention was delivered, and therapeutic orientation) were related to effect size.

Calculations of weighted mean effect sizes, heterogeneity, and moderators were conducted using Comprehensive Meta-Analysis, version 2.2.046 (Borenstein, Hedges, Higgins, & Rothstein, 2005). A decision was made to estimate overall effect sizes using random effects models, as it was presumed that the included studies represent a distribution of true intervention effects. Fixed effect models assume that variability in effect sizes is due to random error within studies, and that there is a common true effect size across all studies. The overall effect size represents the estimate of the true effect size for the population of studies, but is not generalizable beyond the sample of included studies. In contrast, random effects models assume that variability in effect sizes is due to both random error within studies and systematic variability between studies — the true effect size is allowed to vary across studies. Overall effect size represents the estimated average of the true effect sizes, and results can be generalized to studies not included in the analysis. Considerable heterogeneity of effect sizes was expected given the differences in study design, interventions, and samples across the included studies. As the Q statistic is underpowered in cases of small sample size, random effects models were estimated regardless of the observed heterogeneity.

For each of these analyses, presence of outliers was assessed using the sample-adjusted meta-analytic deviance (SAMD) statistic (Huffcutt & Arthur, 1995). A more conservative cutoff score of 2.58 was used to consider studies for exclusion from the analyses, as extreme values can result from either true population variability or error, and removing outliers

whose effects represent true variability limits the ability to assess the role of moderators (Beal, Corey, & Dunlap, 2002). The SAMDs were rank-ordered and the scree plots examined to confirm the outlier status of studies with SAMDs above this cutoff. In cases where the SAMD value was greater than 2.58 but the scree plot suggested that the SAMD was not discrepant from the overall distribution, the study was retained to maximize the variance available to assess the role of moderators.

Publication bias was assessed by visual examination of funnel plots, Duval and Tweedie's (2000) trim-and-fill procedure, and classic fail-safe N values (Rosenthal, 1979). First, the effect size for each study was plotted against the study standard error. An asymmetric distribution suggests missing studies due to publication bias. When asymmetry is present, Duval and Tweedie's (2000) trim-and-fill procedure provides an effect size estimate that corrects for the number and assumed location of the missing studies. When this test indicated significant asymmetry in the funnel plot, the overall estimates for the model were calculated using the trim-and-fill correction. The fail-safe value determines the number of studies with null findings that would be necessary to produce a nonsignificant overall effect size. Using Rosenthal's (1991) recommendation, a value of $5K+10$, where K is the number of observed studies, was used as the cutoff for an unlikely number of studies.

2. Results

2.1. Study characteristics

Table 1 displays characteristics of the 27 studies included in the analyses. Of the included studies, 9 were open trials (33%), 2 were quasi-randomized trials (7%), and 16 were randomized controlled trials (59%). Nineteen studies assessed psychological interventions (70%), four assessed pharmacological interventions (15%), and four assessed interventions including psychological and pharmacological components (15%). Most studies targeted postpartum depression ($n=22$, 81%), four targeted antenatal depression (15%), and one study included subjects across the perinatal period. Length of treatment ranged from 6 to 16 weeks, with an average of 10 weeks. Interpersonal psychotherapy (IPT) was the most common psychological intervention ($n=11$, 41%), followed by cognitive-behavioral (CBT) interventions ($n=9$, 33%); other interventions included nondirective counseling ($n=3$), a Mother-Infant Therapy Group ($n=1$), a CBT-oriented psychoeducational group ($n=1$), manualized supportive psychotherapy ($n=2$), and psychodynamic therapy ($n=2$). Nine studies assessed group interventions (33%), 4 included home-based interventions (15%), and one study included a school-based intervention.

Table 2 presents characteristics of the included studies indicative of their methodological quality. Given the range of designs that were included in the analyses, methodological quality was not quantified or used in the weighting of effect sizes. Of the 27 included studies, 19 included intent-to-treat analyses. As the average attrition rate was 21%, it is likely that completer analyses represent biased outcomes. In 8 studies, subjects were not excluded if they were currently receiving additional treatment for depression, all of which were studies of psychological interventions. In some trials, rates of concurrent antidepressant use were quite high (e.g., 54% across both intervention and control groups, Klier, Muzik, Rosenblum, & Lenz, 2001). While some studies specifically assessed the

potential effect of concurrent antidepressant use on outcome (e.g., Honey, Bennett, & Morgan, 2002), inclusion of subjects receiving adjunctive pharmacological treatment in trials assessing psychological interventions is a significant limitation of the research base for these interventions. Among the 23 studies that included a psychological intervention, most studies provided information regarding therapist characteristics and use of therapy manuals. Fifteen studies provided information about therapist training, 13 indicated that therapists received regular ongoing supervision, and 9 studies assessed for adherence to the treatment model. Among the 8 studies that included a pharmacological intervention, only 3 included a placebo condition in which both patients and clinicians were blind to medication status.

The included studies vary widely in the demographic characteristics and variability of their samples. Most studies were conducted in the United States ($n=13$, 48%), six studies were conducted in Australia (22%), three studies in the United Kingdom (11%), and the remaining studies were conducted in Austria ($n=1$), Canada ($n=1$), France ($n=1$), and Sweden ($n=2$). Fourteen studies (52%) reported at least some information regarding the race, ethnicity, or national origin of subjects. In the 10 studies (37%) that reported demographic information that provided sufficient data regarding the racial composition of their samples, the percentage of subjects who identified as racial/ethnic minorities ranged from 0% (Klier et al., 2001) to 100% (Miller, Gur, Shanok, & Weissman, 2008), with a mean of 45.6% and standard deviation of 32.7%. Twenty-three studies (85%) reported information regarding the marital status of subjects. The percentage of single subjects in these studies ranged from 0% (O'Hara, Stuart, Gorman, & Wenzel, 2000, who required women to be married or living with a partner for 6 months or more to be eligible to participate) to 72.7% (Miller et al., 2008). Twenty studies (74%) reported information regarding parity or the number of children in subjects' households. Of the 14 studies that reported the percentage of primiparous subjects, this value ranged from 25% (O'Hara et al., 2000) to 85.1% (Wiklund, Mohlkert, & Edman, 2010), with a mean of 55.8% and standard deviation of 17.7%.

2.2. Uncontrolled effect sizes

Table 3 presents results of the random effects model for uncontrolled effect sizes, representing results from 25 studies. These values should be interpreted with caution, as they reflect within-study change and cannot differentiate between reductions in symptoms that occurred as a result of the intervention vs. the passage of time. All studies demonstrated significant positive effects, indicating improvement over pretreatment scores, with Hedges' g ranging from 0.78 to 4.39. Two studies had SAMD values greater than 2.58. Visual inspection of the scree plot of the rank-ordered SAMD scores suggested that the value for the study by Grote et al. (2009) was consistent with the overall distribution of SAMD scores, while the study by Appleby, Warner, Whitton, and Faragher (1997) was discrepant. This study was excluded from subsequent analyses; the average effect size excluding this outlier was 1.54 (95% CI 1.34–1.73, $p<0.001$).

The Q statistic indicated that there was significant heterogeneity among the effect sizes ($p<0.001$). The I^2 value indicated a high level of heterogeneity, with 86% of the variance in effect sizes attributable to between-study variance. The fail-safe value was 9102, far exceeding the tolerance level for an unlikely number of non-significant studies (130). The

funnel plot was slightly asymmetric; trim-and-fill procedures suggested that two studies with effect sizes to the right (more strongly positive) of the mean were missing. The corrected average effect size was 1.61 (95% CI 1.40–1.81). This adjusted value suggests that if the included studies do reflect a publication bias, it is in the direction of underestimating the true effect size of the interventions.

As the magnitude of uncontrolled effect sizes is difficult to interpret, a separate meta-analysis was conducted to determine the average level of depressive symptoms at posttreatment. A random effects model was used to calculate the overall mean for the 15 studies that used the EPDS as an outcome measure. The average score at posttreatment was 8.62 (95% CI 7.66–9.58), which is below the commonly used cutoff of 11–13 considered indicative of clinically significant depressive symptoms (Cox, Chapman, Murray, & Jones, 1996). A second random effects model was used to calculate the overall mean for the 5 studies that used the Hamilton Depression Rating Scale as an outcome measure. The average score at posttreatment was 6.12 (95% CI 2.90–9.33), which is also below the cutoff of 7 commonly considered indicative of symptom remission (Frank et al., 1991).

2.3. Moderator analyses: Uncontrolled effect sizes

As both the Q statistic and I^2 index indicated significant heterogeneity of effect sizes, exploratory analyses of potential moderators were conducted. These analyses assessed whether effect sizes differed on the basis of characteristics of the included studies and interventions. As the reporting of sample characteristics was inconsistent across studies, none of these variables was assessed as potential moderators. Subgroups including only one study were excluded from moderator analyses.

2.3.1. Study characteristics—Four characteristics of the included studies were assessed as potential moderators: study design, type of sample, whether a clinician-verified diagnosis of a depressive disorder was required for inclusion in the study, and outcome measure (see Table 4). No significant differences in the average effect size were found among the three types of study designs, between studies assessing interventions for antenatal depression vs. postpartum depression, or between studies that did and did not require a clinician-verified diagnosis for inclusion. Studies for which the effect size was calculated using the BDI had significantly smaller effect sizes ($g=1.10$, $n=2$) than studies for which the effect size was calculated using the EPDS ($g=1.62$, $n=16$) or HDRS ($g=1.65$, $n=5$).

2.3.2. Intervention variables—Two characteristics of the interventions were assessed for potential moderation (see Table 4). Studies including three types of interventions were compared: pharmacological, psychological, and combined (pharmacological +psychological). There were no significant differences in effect sizes among the three major types of interventions. Meta-regression analysis was used to assess the relationship between length of treatment and effect size. There was a trend for a positive association of length of treatment with effect size; however, this result did not reach significance (slope=0.03, $p=0.07$).

For studies that included a psychological intervention, three characteristics of the intervention were assessed for moderation: method of administration (individual vs. group),

location of administration (clinic vs. home), and therapeutic orientation. Studies in which therapy was administered individually had significantly larger effect sizes ($g=1.79$, $n=12$) than those utilizing a group therapy format ($g=1.23$, $n=7$). There was a trend ($p=0.08$) toward home-administered treatments ($g=2.33$, $n=2$) having larger effect sizes than clinic-based treatments ($g=1.64$, $n=16$).

Four analyses were conducted to assess whether inclusion of two well-established therapeutic interventions, cognitive-behavioral therapy (CBT) or interpersonal psychotherapy (IPT), was a moderator of effect size. First, studies were categorized as either including or not including each approach. There was a trend for studies that included one of these interventions having larger effect sizes ($g=1.61$, $n=18$) than studies that included other psychological interventions ($g=1.11$, $n=2$). There were not significant differences between studies that included CBT compared to those that did not (including IPT interventions) or between studies that included IPT compared to those that did not (including CBT interventions). Finally, the effect sizes of studies including CBT and IPT were compared to one another. No study included both a CBT and an IPT intervention. There was not a significant difference in the average effect size of the two interventions.

2.4. Controlled effect sizes

Table 5 presents the results of the random effects model for controlled effect sizes, representing results from 14 studies. All effect sizes were positive, indicating superiority of treatment to control conditions, with Hedges' g ranging from 0.31 to 2.33. Two studies had SAMD values greater than 2.58. Visual inspection of the scree plot of the rank-ordered SAMD scores indicated that the value for the study by Milgrom, Negri, Gemmill, McNeil, and Martin (2005) was consistent with the overall distribution of SAMD scores, while the study by Chabrol et al. (2002) was discrepant. This study was excluded from subsequent analyses. The average effect size, excluding the outlier, was 0.65 (95% CI 0.45–0.86, $p<0.001$). Cohen's $U3$ metric provides an intuitive metric through which to interpret the magnitude of this effect size; this value indicates that 74% of subjects in treatment conditions could be expected to report levels of depressive symptoms lower than the mean of the control group (Lipsey & Wilson, 2001).

The Q statistic indicated that there was significant heterogeneity among the effect sizes ($p<0.05$). The I^2 value indicated a medium level of heterogeneity, with 43% of the variance in effect sizes attributable to between-study variance. The fail-safe N was 229, which substantially exceeds the tolerance level for an unlikely number of nonsignificant studies (75). The funnel plot was symmetric and the trim-and-fill procedures suggested no missing studies.

2.5. Moderator analyses: Controlled effect sizes

As both the Q statistic and I^2 value indicated heterogeneity among effect sizes, exploratory analyses of potential moderators were conducted to assess whether effect sizes differed on the basis of study or intervention characteristics. Subgroups including only one study were excluded from moderator analyses.

2.5.1. Study characteristics—As for uncontrolled effect sizes, target population, diagnostic status, and outcome measure were assessed as potential moderators (see Table 6). In addition, the average effect size for studies utilizing different control groups (treatment as usual vs. enhanced treatment as usual vs. waiting list control) was assessed. The only significant moderator of effect size was the target population of the study; interventions for antenatal depression had significantly larger effect sizes ($g=1.18$, $n=2$) than those for postpartum depression ($g=0.57$, $n=11$).

2.5.2. Intervention variables—As only one study assessing pharmacological treatment was included in this analysis, it was not possible to compare the average effect size for pharmacological vs. psychological interventions. Length of treatment was assessed for moderation using meta-regression analysis; treatment length was not significantly associated with effect size (slope=0.05, $p=0.12$).

For the studies assessing psychological interventions, three potential moderators were assessed (see Table 6). There were no significant differences among average effect sizes with respect to the method of administration of therapy (individual vs. group vs. combined). There was a trend for studies that included clinic-based interventions to have larger effect sizes ($g=0.73$, $n=9$) than studies that included home-based interventions ($g=0.38$, $n=2$). Studies that included an IPT intervention were compared to those that included a CBT intervention. As only one study included an intervention representing different therapeutic orientation, analyses comparing IPT and CBT separately to all other treatments combined were not conducted. Studies that included an IPT intervention had significantly larger effect sizes ($g=0.96$, $n=5$) than those that included a CBT intervention ($g=0.40$, $n=6$).

3. Discussion

Results of these analyses provide evidence for the efficacy of a range of interventions for perinatal depression. All studied interventions demonstrated symptomatic improvement from pretreatment to posttreatment, with posttreatment means for both the EPDS and HDRS below cutoffs for clinically significant depressive symptoms. All interventions also demonstrated superiority to control conditions, with an overall effect size in the moderate range. The overall effect size ($g=0.65$) is comparable to that found in a meta-analysis of the efficacy of psychological treatments for adult depression, which reported an average posttreatment effect size of 0.67 (Cuijpers, Smit, Bohlmeijer, Hollon, & Andersson, 2010).

Our finding indicating the superiority of IPT to CBT for the treatment of perinatal depression has important implications for both clinicians and researchers. This is consistent with the findings of Bledsoe and Grote (2006), who reported that IPT interventions were associated with greater decreases in symptoms from pretreatment to posttreatment; our study is the first to find that IPT results in a greater reduction in symptoms compared to control conditions, as well. It is possible that this finding reflects a true difference in the efficacy of a specific form of psychotherapy for this population. In describing their adaptations of IPT for postpartum depression, Stuart and O'Hara (1995) noted that the focus on interpersonal problem areas, particularly role transitions and interpersonal disputes, may be particularly well-suited to the problems women experience during the perinatal period, such as

disruptions in their interpersonal relationships. While a possible interpretation of this finding is that IPT is a more efficacious intervention for depression among this population, our findings may also have resulted from characteristics of the included studies unrelated to the interventions themselves. Studies assessing an IPT intervention were more likely to report utilization of a therapy manual; the implementation of manualized IPT for this population is likely facilitated by the ready availability of a treatment manual containing specific adaptations for postpartum depression (O'Hara et al., 2000). It is unclear whether studies of CBT interventions did not utilize a specific therapy manual or whether these studies simply followed widely accepted and available manuals for CBT for depression (e.g., Beck, Rush, Shaw, & Emery, 1979). Studies assessing a CBT intervention were also more likely to include other interventions, either as an explicit element of the intervention itself (e.g., Honey et al., 2002, whose intervention included educational, cognitive-behavioral, and relaxation components), or because a single effect size was calculated for studies including multiple active treatments for the purpose of our analyses (e.g., Cooper, Murray, Wilson, & Romaniuk, 2003, who included CBT, psychodynamic, and nondirective counseling interventions). Given the possibility that other aspects of study design may have been confounded with the therapeutic orientation of the interventions included in the different studies, our results regarding the superiority of IPT to CBT should be interpreted with caution. Further research evaluating the efficacy of well-defined cognitive-behavioral interventions for this population, particularly trials in which faithfully administered CBT and IPT protocols are compared directly to one another, is necessary to establish whether IPT is truly a more effective intervention for perinatal depression. Such studies would be greatly beneficial given the low acceptability of antidepressant medication in this population and preference for non-pharmacologic treatments, including psychotherapy (Kim et al., 2011).

Another important moderator identified in these analyses is the superiority of individually-administered therapeutic interventions to those conducted in a group format. Uncontrolled effect sizes were larger for studies in which therapy was conducted on an individual basis; a comparable pattern was observed for controlled effect sizes, although this difference did not reach significance due to the smaller number of included studies. There has been great interest in the potential use of group treatments for perinatal depression, and some have suggested that the format may be particularly well-suited for this population because it provides an opportunity for normalizing women's experiences and decreasing social isolation (Mulcahy, Reay, Wilkinson, & Owen, 2009). Results of these analyses suggest that these benefits may not be sufficient to lead to a reduction in depressive symptoms comparable to that which can be achieved through individual psychotherapy.

Our results suggest that further assessment of home-based interventions is necessary in order to determine the efficacy of these programs. There was a trend for depressive symptoms to decrease more over time in studies that incorporated home-based interventions; however, there was also a trend for greater effects in studies that incorporated clinic-based interventions with respect to the superiority of treated groups to control conditions. Both of these results should be interpreted with caution, as only two studies of home-based interventions were included in these moderator analyses. Further research, ideally trials in which comparable interventions administered either in a clinic or home setting can be

directly compared, is necessary to determine whether the location in which therapy is conducted affects the efficacy of these interventions.

There was a trend for longer treatments to have larger uncontrolled effect sizes, but treatment length was not associated with controlled effect sizes. As uncontrolled effect sizes do not distinguish between the effects of interventions and decreases in symptoms over time, this raises the concern that the apparent effects of treatment may simply reflect a natural decrease in symptoms over time. There is evidence that, for most women, high levels of depressive symptoms naturally remit over the perinatal period (Heron et al., 2004). One study included in these analyses found that three active treatment groups were only superior to routine primary care at 4.5 months postpartum; by 9 months postpartum the depressive symptoms of the control group had declined to levels comparable to those observed in the treated groups, and differences among the groups remained nonsignificant through five years postpartum (Cooper et al., 2003). However, other included studies found that treatment groups continued to be superior to control groups at follow-up times ranging from 3 to 6 months posttreatment (Grote et al., 2009; Honey et al., 2002; Mulcahy et al., 2009).

Unfortunately, the number of studies that has included follow-up assessments of treated subjects is too small for a meta-analysis of long-term outcomes. Five studies in which an intervention was compared to a control group reported follow-up outcomes, with the timing of follow-up assessments ranging from 3 months to 5 years posttreatment. An initial analysis suggested that the treated group remained superior at the first posttreatment follow-up assessment (which ranged from 3 to 6 months posttreatment), with a positive effect size in the moderate range. However, the SAMD values indicated that two studies were significant outliers, and the fail-safe N for this analysis indicated that the number of studies with null results necessary to reduce the effect size to zero was below the tolerance limit. Further assessment of long-term outcomes for subjects in controlled trials of interventions for perinatal depression are necessary to assess whether the benefits of treatment for perinatal depression are maintained over time.

The present meta-analysis has several limitations. Reflecting the status of the field, the number of included studies is relatively small, particularly for analysis of controlled effect sizes. Moderator analyses were likewise limited by the small number of included studies. The quality of studies assessing interventions among this population is somewhat limited (see also Cuijpers, Brännmark, & van Straten, 2008; Dennis, 2004). Of particular concern is the fact that 8 studies of psychological interventions did not exclude subjects who were receiving pharmacological treatment, and rates of concurrent antidepressant use were quite high in some of these studies. As this raises the possibility that the purported effects of psychotherapy in these studies could have been the result of pharmacological interventions, future research is necessary both to establish the separate effects of psychotherapy and pharmacotherapy (by excluding subjects receiving concurrent treatment from research) and to explicitly assess the efficacy of combined psychological and pharmacological treatments. Compared with psychological interventions, there has been relatively little assessment of the efficacy of antidepressant medication in this population. Of the seven studies that included antidepressant treatment, three were open trials and two compared a combined treatment to either psychological or pharmacological monotherapy. Given the ethical concerns regarding

the use of no-treatment control groups in treatment studies with this population, further studies in which pharmacological treatments are directly compared to psychological or combined interventions are necessary to address the relative efficacy of these interventions. Barber (2009) also suggested that there is room for large scale, relatively well controlled, naturalistic studies for examining the efficacy of psychotherapy, as the field does not have the resources to conduct all the RCTs that need to be conducted. Finally, because psychotherapies are large packages, research that focuses on specific interventions from a package could be used to determine which specific interventions are particularly useful (e.g., Barber & Crits-Christoph, 1996; DeRubeis & Feeley, 1990; Webb, DeRubeis, & Barber, 2010).

In summary, these meta-analyses demonstrated that a range of interventions is effective in the reduction of perinatal depressive symptoms. Reductions in symptoms from prettest to posttreatment are large, and symptom levels at posttreatment are below cutoffs for clinically significant symptoms. These interventions reliably lead to moderate reductions in depressive symptoms compared to control groups. Interestingly, there was initial evidence that IPT may be more effective than CBT, although further research is necessary to establish whether this can be attributed to methodological differences between studies assessing the two forms of psychotherapy. Relatively few studies of antidepressant medication for this population have been conducted compared to psychological interventions, and overall most interventions have not been assessed in comparison to control or other active treatment conditions. Given the prevalence of perinatal depression and the negative outcomes associated with depressive symptoms during this period, the identification of effective and acceptable treatments for this population is vitally necessary. Although more research is needed to confirm and extend the results of these meta-analyses, these results suggest a range of interventions for further investigation as treatments for this disorder.

References

References marked with an asterisk indicate studies included in the meta-analyses. References marked with two asterisks indicate secondary sources and were only used when insufficient outcome data was presented in the primary article.

- *. Appleby L, Warner R, Whitton A, Faragher B. A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. *British Medical Journal*. 1997; 314:932–936. [PubMed: 9099116]
- Barber JP. Towards a working through of some core conflicts in psychotherapy research. *Psychotherapy Research*. 2009; 19:1–12. [PubMed: 19206018]
- Barber JP, Crits-Christoph P. Development of an adherence/competence scale for dynamic therapy: Preliminary findings. *Psychotherapy Research*. 1996; 6:81–94. [PubMed: 22242608]
- Beal DJ, Corey DM, Dunlap WP. On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in the meta-analysis of correlations. *Journal of Applied Psychology*. 2002; 87:583–589. [PubMed: 12090616]
- Beck, AT.; Rush, AJ.; Shaw, BF.; Emery, G. *Cognitive therapy of depression*. New York: Guilford Press; 1979.
- Bennett HA, Einarson A, Taddio A, Koren G, Einarson TR. Depression during pregnancy: Overview of clinical factors. *Clinical Drug Investigation*. 2004a; 24:157–179. [PubMed: 17516702]

- Bennett HA, Einarson A, Taddio A, Koren G, Einarson TR. Prevalence of depression during pregnancy: Systematic review. *Obstetrics & Gynecology*. 2004b; 103:698–709. [PubMed: 15051562]
- Bledsoe SE, Grote NK. Treating depression during pregnancy and the postpartum: A preliminary meta-analysis. *Research on Social Work Practice*. 2006; 16(2):109–120.
- Borenstein, M.; Hedges, L.; Higgins, J.; Rothstein, H. *Comprehensive meta-analysis version 2*. Englewood NJ: Biostat; 2005.
- Chabrol H, Teissedre F, Armitage J, Danel M, Walburg V. Acceptability of psychotherapy and antidepressants for postnatal depression among newly delivered mothers. *Journal of Reproductive and Infant Psychology*. 2004; 22:5–12.
- *. Chabrol H, Teissedre F, Saint-Jean M, Teisseyre N, Rogé B, Mullet E. Prevention and treatment of post-partum depression: A controlled randomized study on women at risk. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences*. 2002; 32(6):1039–1047.
- *. Clark R, Tluczek A, Wenzel A. Psychotherapy for postpartum depression: A preliminary report. *American Journal of Orthopsychiatry*. 2003; 73:441–454. [PubMed: 14609406]
- *. Cohen LS, Viguera AC, Bouffard SM, Nonacs RM, Morabito C, Collins MH, et al. Venlafaxine in the treatment of postpartum depression. *Journal of Clinical Psychiatry*. 2001; 62(8):592–596. [PubMed: 11561929]
- Cooper PJ, Murray L. Course and recurrence of postpartum depression: Evidence for the specificity of the diagnostic concept. *British Journal of Psychiatry*. 1995; 166:191–195. [PubMed: 7728362]
- **. Cooper, PJ.; Murray, L. The impact of psychological treatments of postpartum depression on maternal mood and infant development. In: Murray, L.; Cooper, P., editors. *Postpartum depression and child development*. New York: Guilford Press; 1997. p. 201–220.
- *. Cooper PJ, Murray L, Wilson A, Romaniuk H. Controlled trial of the short-and long-term effect of psychological treatment of post-partum depression 1: Impact on maternal mood. *British Journal of Psychiatry*. 2003; 182(5):412–419. [PubMed: 12724244]
- Cox JL, Chapman G, Murray D, Jones P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. *Journal of Affective Disorders*. 1996; 39:185–189. [PubMed: 8856422]
- *. Craig E, Judd F, Hodgins G. Therapeutic group programme for women with postnatal depression in rural Victoria: A pilot study. *Australasian Psychiatry*. 2005; 13:291–295. [PubMed: 16174204]
- Cuijpers P, Brännmärk JG, van Straten A. Psychological treatment of postpartum depression: A meta-analysis. *Journal of Clinical Psychology*. 2008; 64:103–118. [PubMed: 18161036]
- Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: Meta-analytic study of publication bias. *British Journal of Psychiatry*. 2010; 196:173–178. [PubMed: 20194536]
- Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology*. 2008; 76:909–922. [PubMed: 19045960]
- Dawes M, Chowienzyk PJ. Pharmacokinetics in pregnancy. *Best Practice & Research Clinical Obstetrics & Gynaecology*. 2001; 15:819–826. [PubMed: 11800526]
- Dennis CL. Treatment of postpartum depression, part 2: A critical review of nonbiological interventions. *Journal of Clinical Psychiatry*. 2004; 65:1252–1265. [PubMed: 15367054]
- Dennis CL, Hodnett ED. Psychosocial and psychological interventions for treating postpartum depression. *Cochrane Database of Systematic Reviews*. 2007
- Dennis CL, Ross L. Relationships among infant sleep patterns, maternal fatigue, and development of depressive symptomatology. *Birth*. 2005; 32:187–193. [PubMed: 16128972]
- DeRubeis RJ, Feeley M. Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research*. 1990; 14:469–482.
- Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000; 56:455–463. [PubMed: 10877304]
- Epperson CN. Postpartum major depression: Detection and treatment. *American Family Physician*. 1999; 59:2247–2254. [PubMed: 10221309]

- Epperson CN, Anderson GM, McDougale CJ. Sertraline and breast-feeding. *New England Journal of Medicine*. 1997; 336:1189–1190. [PubMed: 9102576]
- Epperson CN, Jatlow PI, Czarkowski K, Anderson GM. Maternal fluoxetine treatment in the postpartum period: Effects on platelet serotonin and plasma drug levels in breastfeeding mother–infant pairs. *Pediatrics*. 2003; 112:425.
- Feske U, Chambless DL. Cognitive behavioral versus exposure only treatment for social phobia: A meta-analysis. *Behavior Therapy*. 1995; 26:695–720.
- Forray A, Mayes LC, Magriples U, Epperson CN. Prevalence of post-traumatic stress disorder in pregnant women with prior pregnancy complications. *Journal of Maternal and Fetal Neonatal Medicine*. 2009; 22:522–527.
- Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Archives of General Psychiatry*. 1991; 48:851–855. [PubMed: 1929776]
- Freeman MP, Davis M, Sinha P, Wisner KL, Hibbeln JR, Gelenberg AJ. Omega-3 fatty acids and supportive psychotherapy for perinatal depression: A randomized placebo-controlled study. *Journal of Affective Disorders*. 2008; 110:142–148. [PubMed: 18206247]
- Gavin NI, Gayner BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: A systematic review of prevalence and incidence. *Obstetrics & Gynecology*. 2005; 106(5):1071–1083. [PubMed: 16260528]
- **. Goodman SH, Broth MR, Hall CM, Stowe ZN. Treatment of postpartum depression in mothers: Secondary benefits to the infants. *Infant Mental Health Journal*. 2008; 29:492–513.
- Grace SL, Evindar A, Stewart DE. The effect of postpartum depression on child cognitive development and behavior: A review and critical analysis of the literature. *Archives of Women's Mental Health*. 2003; 6:263–274.
- Grote NK, Bridge JA, Gavin AR, Melville JL, Iyengar S, Katon WJ. A meta-analysis of depression during pregnancy and the risk of preterm birth, low birth weight, and intrauterine growth restriction. *Archives of General Psychiatry*. 2010; 67(10):1012–1024. [PubMed: 20921117]
- *. Grote NK, Swartz HA, Geibel SL, Zuckoff A, Houck PR, Frank E. A randomized controlled trial of culturally relevant, brief interpersonal psychotherapy for perinatal depression. *Psychiatric Services*. 2009; 60:313–321. [PubMed: 19252043]
- Halbreich U, Karkun S. Cross-cultural and social diversity of prevalence of postpartum depression and depressive symptoms. *Journal of Affective Disorders*. 2006; 91:97–111. [PubMed: 16466664]
- Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*. 1981; 6:107–128.
- Heron J, O'Connor TG, Evans J, Golding J, Glover V. The course of anxiety and depression through pregnancy and the postpartum in a community sample. *Journal of Affective Disorders*. 2004; 80:65–73. [PubMed: 15094259]
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *British Medical Journal*. 2003; 327:557–560. [PubMed: 12958120]
- *. Honey KL, Bennett P, Morgan M. A brief psycho-educational group intervention for postnatal depression. *British Journal of Clinical Psychology*. 2002; 41:405–409. [PubMed: 12437794]
- Hostetter A, Stowe ZN, Strader JR, McLaughlin E, Llewellyn A. Dose of selective serotonin uptake inhibitors across pregnancy: Clinical implications. *Depression & Anxiety*. 2000; 11:51–57. [PubMed: 10812529]
- Huffcutt AI, Arthur W Jr. Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*. 1995; 80:327–334.
- Joffe R, Sokolov S, Streiner D. Antidepressant treatment of depression: A meta-analysis. *Canadian Journal of Psychiatry*. 1996; 41:613–616.
- Kammerer M, Taylor A, Glover V. The HPA axis and perinatal depression: A hypothesis. *Archives of Women's Mental Health*. 2006; 9:187–196.
- Kim DR, O'Reardon JP, Epperson CN. Guidelines for the management of depression during pregnancy. *Current Psychiatry Reports*. 2010; 12:279–281. [PubMed: 20424977]

- Kim DR, Sokol L, Barber JP, Moseley M, Lamprou L, O'Reardon JP, et al. A survey of patient acceptability of repetitive Transcranial Magnetic Stimulation (TMS) during pregnancy. *Journal of Affective Disorders*. 2011; 129:385–390. [PubMed: 20864179]
- *. Klier CM, Muzik M, Rosenblum KL, Lenz G. Interpersonal psychotherapy adapted for the group setting in the treatment of postpartum depression. *Journal of Psychotherapy Practice and Research*. 2001; 10:124–131. [PubMed: 11264336]
- Lipsey, MW.; Wilson, D. *Applied Social Research Methods Series*. Vol. 49. Thousand Oaks, CA: SAGE Publications, Inc; 2001. Practical meta-analysis.
- *. Meager I, Milgrom J. Group treatment for postpartum depression: A pilot study. *Australian and New Zealand Journal of Psychiatry*. 1996; 30(6):852–860. [PubMed: 9034477]
- *. Milgrom J, Negri LM, Gemmill AW, McNeil M, Martin PR. A randomized controlled trial of psychological interventions for postnatal depression. *British Journal of Clinical Psychology*. 2005; 44(4):529–542. [PubMed: 16368032]
- *. Miller L, Gur M, Shanok A, Weissman M. Interpersonal psychotherapy with pregnant adolescents: Two pilot studies. *Journal of Child Psychology and Psychiatry*. 2008; 49(7):733–742. [PubMed: 18492043]
- *. Misri S, Reebye P, Corral M, Mills L. The use of paroxetine and cognitive-behavioral therapy in postpartum depression and anxiety: A randomized controlled trial. *Journal of Clinical Psychiatry*. 2004; 65(9):1236–1241. [PubMed: 15367052]
- *. Mulcahy R, Reay RE, Wilkinson RB, Owen C. A randomised controlled trial for the effectiveness of group interpersonal psychotherapy for postnatal depression. *Archives of Women's Mental Health*. 2009; 13:125–139.
- **. Murray L, Cooper PJ, Wilson A, Romaniuk H. Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression 2: Impact on the mother–child relationship and child outcome. *British Journal of Psychiatry*. 2003; 182:420–427. [PubMed: 12724245]
- O'Hara MW, Swain AM. Rates and risk of postpartum depression: A metaanalysis. *International Review of Psychiatry*. 1996; 8:37–54.
- *. O'Hara MW, Stuart S, Gorman LL, Wenzel A. Efficacy of interpersonal psychotherapy for postpartum depression. *Archives of General Psychiatry*. 2000; 57:1039–1045. [PubMed: 11074869]
- *. Pearlstein TB, Zlotnick C, Battle CL, Stuart S, O'Hara MW, Price AB, et al. Patient choice of treatment for postpartum depression: A pilot study. *Archives of Women's Mental Health*. 2006; 9:303–308.
- *. Prendergast J, Austin MP. Early childhood nurse-delivered cognitive behavioural counselling for post-natal depression. *Australasian Psychiatry*. 2001; 9(3):255–259.
- *. Reay R, Fisher Y, Robertson M, Adams E, Owen C. Group interpersonal psychotherapy for postnatal depression: A pilot study. *Archives of Women's Mental Health*. 2006; 9:31–39.
- Rosenthal R. The “file drawer problem” and tolerance for null results. *Psychological Bulletin*. 1979; 86:638–641.
- Rosenthal, R. *Meta-analytic procedures for social research*. Newbury Park, CA: Sage; 1991.
- *. Spinelli MG. Interpersonal psychotherapy for depressed antepartum women: A pilot study. *American Journal of Psychiatry*. 1997; 154:1028–1030. [PubMed: 9210760]
- *. Spinelli MG, Endicott J. Controlled clinical trial of interpersonal psychotherapy versus parenting education program for depressed pregnant women. *The American Journal of Psychiatry*. 2003; 160(3):555–562. [PubMed: 12611838]
- *. Stowe ZN, Casarella J, Landrey J, Nemeroff CB. Sertraline in the treatment of women with postpartum major depression. *Depression*. 1995; 3:49–55.
- *. Stuart S, O'Hara MW. Interpersonal therapy for postpartum depression: A treatment program. *Journal of Psychotherapy Practice and Research*. 1995; 4:18–29. [PubMed: 22700210]
- *. Suri R, Burt VK, Altshuler LL. Nefazodone for the treatment of postpartum depression. *Archives of Women's Mental Health*. 2005; 8(1):55–56.
- Webb CA, DeRubeis RJ, Barber JP. Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*. 2010; 78:200–211. [PubMed: 20350031]

- *. Wickberg B, Hwang CP. Counselling of postnatal depression; a controlled study on a population based Swedish sample. *Journal of Affective Disorders*. 1996; 39:209–216. [PubMed: 8856425]
  - *. Wiklund I, Mohlkert P, Edman G. Evaluation of a brief cognitive intervention in patients with signs of postnatal depression: A randomized controlled trial. *Acta Obstetrica et Gynecologica*. 2010; 89:1100–1104.
- Wisner KL, Perel JM, Wheeler SB. Tricyclic dose requirements across pregnancy. *American Journal of Psychiatry*. 1993; 150:1541–1542. [PubMed: 8379562]
- *. Yonkers KA, Lin H, Howell HB, Health AC, Cohen LS. Pharmacologic treatment of postpartum women with new-onset major depressive disorder: A randomized controlled trial with paroxetine. *Journal of Clinical Psychiatry*. 2008; 69:659–665. [PubMed: 18363420]

Fig. 1.
Flow chart illustrating the identification of included studies.

Table 1

Characteristics of studies assessing interventions for perinatal depression.

Study	Country	N	Study design	Control type	Population	Intervention	Dx	Measure	Tx Length	% Attrition	Treatment	Format	Admin
Appleby et al. (1997)	UK	87	RCT		POST	COMB	Y	EPDS	12	30	CBT	IND	CLIN
Chabrol et al. (2002)	FR	48	RCT	TAU	POST	THER	Y	EPDS	12	0	CBT	IND	HOME
Clark, Tluczek, and Wenzel (2003)	US	39	QRT	WL	POST	THER	Y	BDI	12	10	M-ITG+IPT	COMB	CLIN
Cohen et al. (2001)	US	15	OT		POST	MED	Y	HDRS	8	33	Venlafaxine		
Cooper et al. (2003)	UK	193	RCT	TAU	POST	THER	Y	EPDS	10	17	CBT+NDC+PDT	IND	HOME
Craig, Judd, and Hodgins (2005)	AUS	16	OT		POST	THER	N	EPDS	9	13	CBT	GRP	CLIN
Freeman, Davis, Sinha, Wisner, Hibbeln, and Gelenberg (2008)	US	59	RCT		MIXED	COMB	Y	EPDS	8	34	SUPP+omega-3 fatty acids	IND	IND
Grote et al. (2009)	US	53	RCT	TAU+	ANTE	THER	Y	EPDS	NS	13	IPT	IND	CLIN
Honey et al. (2002)	UK	45	RCT	TAU	POST	THER	N	EPDS	8	9	PEG	GRP	CLIN
Klier et al. (2001)	AUST	17	OT		POST	THER	Y	EPDS	12	35	IPT	GRP	CLIN
Meager and Milgrom (1996)	AUS	20	RCT	WL	POST	THER	N	EPDS	10	40	CBT	GRP	CLIN
Milgrom et al. (2005)	AUS	120	RCT	TAU	POST	THER	Y	BDI	12	37	CBT+NDC	COMB	CLIN
Miller et al. (2008)	US	11	OT		ANTE	THER	Y	EPDS	12	0	IPT	GRP	SCHOOL
Misri, Reebye, Corral, and Mills (2004)	CAN	35	RCT		POST	COMB	Y	EPDS	12	9	CBT	IND	CLIN
Mulcahy et al. (2009)	AUS	50	RCT	TAU	POST	THER	Y	EPDS	8	15	IPT	GRP	CLIN
O'Hara et al. (2000)	US	99	RCT	WL	POST	THER	Y	HDRS	12	18	IPT	IND	CLIN
Pearlstein et al. (2006)	US	23	QRT		POST	COMB	Y	EPDS	12	22	IPT	IND	CLIN
Prendergast and Austin (2001)	AUS	37	RCT	TAU+	POST	THER	Y	EPDS	6	0	CBT	IND	HOME
Reay, Fisher, Robertson, Adams, and Owen (2006)	AUS	18	OT		POST	THER	Y	EPDS	8	6	IPT	GRP	CLIN
Spinelli (1997)	US	13	OT		ANTE	THER	Y	EPDS	16	31	IPT	IND	CLIN
Spinelli and Endicott (2003)	US	50	RCT	ACT	ANTE	THER	Y	EPDS	16	24	IPT	IND	CLIN
Stowe, Casarella, Landrey, and Nemeroff (1995)	US	26	OT		POST	MED	Y	BDI	8	19	Sertraline		
Stuart and O'Hara (1995)	US	12	OT		POST	THER	Y	HDRS	NS	42	IPT	IND	CLIN
Suri, Burt, and Alshuler (2005)	US	4	OT		POST	MED	Y	HDRS	8	25	Nefazodone		
Wickberg and Hwang (1996)	SWE	48	RCT	TAU	POST	THER	N	MADRS	6	15	NDC	IND	COMB
Wiklund et al. (2010)	SWE	66	RCT	TAU	POST	THER	N	EPDS	7	0	CBT	IND	CLIN

Study	Country	N	Study design	Control type	Population	Intervention	Dx	Measure	Tx Length	% Attrition	Treatment	Format	Admin
Yonkers, Lin, Howell, Health, and Cohen (2008)	US	70	RCT	PLA	POST	MED	Y	HDRS	8	56	Paroxetine		

Note. Dx = required clinician-administered diagnostic assessment, Tx Length = treatment length in weeks, Admin = location of therapy administration, AUS = Australia, AUST = Austria, CAN=Canada, FR = France, SWE= Sweden, UK = United Kingdom, US = United States, OT=open trial, QRT=quasi-randomized trial, RCT = randomized controlled trial, TAU = treatment as usual, TAU+ = enhanced treatment as usual, WL = waiting list, ACT = active control, PLA = placebo, ANTE = antepartum, POST = postpartum, MIXED = antepartum + postpartum, MED = antidepressant medication, COMB = combined antidepressant medication + psychotherapy, THER = psychotherapy, BDI = Beck Depression Inventory, EPDS = Edinburgh Postnatal Depression Scale, HDRS = Hamilton Depression Rating Scale, MADRS = Montgomery–Asberg Depression Rating Scale, CBT = cognitive–behavioral therapy, IPT = interpersonal psychotherapy, M–ITG = mother–infant therapy group, NDC = nondirective counseling, PEG = psychoeducational group, PDT = psychodynamic therapy, SUPP = manualized supportive psychotherapy, COMB = combined individually + group administered, IND = individually-administered, GRP = group-administered, CLIN = clinic-based intervention, HOME = home-based intervention, COMB = combined clinic+home-based intervention, and SCHOOL = school-based intervention.

Table 2

Methodological quality of studies assessing interventions for perinatal depression.

Study	ITT	Char Sample	Concurr Tx	Blind Assess	RCTs		Psychological interventions				Pharmacologic interventions		
					Random	Spec Ther	Manual	Trng	Super	Adher	Blind Clinic	Blind Pt	
Open trials													
Cohen et al. (2001)	+	+	NS	NS								-	-
Craig et al. (2005)	-	-	NS	NS		+	+	+	-	-			
Klier et al. (2001)	+	+	Y	-		+	+	+	-	-			
Miller et al. (2008)	+	+	N	+		+	+	-	-	-			
Reay et al. (2006 US)	+	+	Y	+		+	+	+	+	+			
Spinelli (1997)	+	+	NS	NS		-	-	-	-	-			
Stowe et al. (1995)	-	+	NS	-								-	-
Stuart and O'Hara (1995)	-	+	Y	NS		-	-	-	-	-			
Suri et al. (2005)	-	-	NS	+								-	-
Quasi-randomized trials													
Clark et al. (2003)	-	+	Y	NS		+	+	+	+	-			
Pearlstein et al. (2006)	-	+	N	NS		+	-	+	+	-		-	-
Randomized controlled trials													
Appleby et al. (1997)	+	+	NS	+	+	+	-	-	+	-	+	+	+
Chabrol et al. (2002)	+	+	N	-	-	+	+	+	+	+			
Cooper et al. (2003)	+	+	NS	+	-	+	+	+	+	+			
Freeman et al. (2008)	+	+	N	NS	-	-	+	-	-	-	+	+	+
Grote et al. (2009)	+	+	N	NS	+	+	+	+	+	+			
Honey et al. (2002)	+	+	Y	NA	+	+	-	-	-	-			
Meager and Milgrom (1996)	-	+	Y	NA	-	+	-	-	-	-			
Milgrom et al. (2005)	+	+	N	NA	-	+	+	+	+	+			
Misri et al. (2004)	+	+	N	-	+	+	+	-	-	-	-	-	-
Mulcahy et al. (2009)	+	+	Y	+	+	+	+	+	+	+			
O'Hara et al. (2000)	+	+	NS	-	+	+	+	+	+	+			
Prendergast and Austin (2001)	+		Y	NS	+	+	+	+	+	+	+	+	+
Spinelli and Endicott (2003)	+	+	N	NS	+	+	+	+	+	+	+	+	+

Study	ITT	Char Sample	Concurr Tx	Blind Assess	RCTs		Psychological interventions				Pharmacologic interventions	
					Random	Spec Ther	Manual	Trng	Super	Adher	Blind Clinic	Blind Pt
Wickberg and Hwang (1996)	-	+	N	+	+	+	-	+	+	-		
Wiklund et al. (2010)	+	+	Y	NA	+	+	-	-	-	-		
Yonkers et al. (2008)	+	+	N	+	+						+	+

Note. ITT = report intent-to-treat analyses, Char Sample = specify characteristics of sample, Concurr Tx = subjects allowed to receive concurrent antidepressant or psychological treatment, Blind Assess = clinician-administered diagnostic measures conducted by independent evaluator blind to treatment condition, Random = specification of method of randomization, Spec Ther = specify therapist characteristics, Manual = specify use of therapy manual, Trng = describe therapist training, Super = indicate therapist supervision, Adher = indicate adherence to model, Blind Clinic = clinician blind to treatment status, Blind Pt = patient blind to treatment status, + = yes, - = no, NA = not applicable, NS = not specified, Y = yes, and N = no.

Table 3
Random weighted uncontrolled effect sizes from studies assessing interventions for perinatal depression.

Study	<i>n</i>	Hedges' <i>g</i>	SAMD
Appleby et al. (1997)	87	4.39***	10.70 ^d
Chabrol et al. (2002)	18	1.92***	0.46
Clark et al. (2003)	24	1.03***	-1.30
Cohen et al. (2001)	15	1.68***	0.03
Craig et al. (2005)	14	1.44***	-0.33
Freeman et al. (2008)	51	1.29***	-1.12
Grote et al. (2009)	25	3.00***	2.82
Honey et al. (2002)	23	0.78***	-1.76
Klier et al. (2001)	17	1.35***	-0.52
Meager and Milgrom (1996)	6	0.95*	-0.59
Miller et al. (2008)	11	0.91***	-0.98
Misri et al. (2004)	35	1.72***	0.17
Mulcahy et al. (2009)	23	1.56***	-0.20
O'Hara et al. (2000)	48	2.05***	1.15
Pearlstein et al. (2006)	23	2.66***	2.02
Prendergast and Austin (2001)	17	2.60***	1.61
Reay et al. (2006)	18	1.61***	-0.09
Spinelli and Endicott (2003)	21	1.20***	-0.88
Spinelli (1997)	13	1.28***	-0.55
Stowe et al. (1995)	19	1.20***	-0.83
Stuart and O'Hara (1995)	6	1.39***	-0.23
Suri et al. (2005)	3	1.67***	0.00
Wickberg and Hwang (1996)	20	0.83***	-1.53
Wiklund et al. (2010)	66	2.03***	-0.40

Study	<i>n</i>		Hedges' <i>g</i>	SAMD
Yonkers et al. (2008)	17	1.42***	0.92	
Total (all studies)	25	1.66***	1.41–1.91	280.99(24)***
Total (outlier excluded)	24	1.54***	1.34–1.73	159.22(23)***
Total (trim-and-fill correction)		1.61	1.40–1.81	199.09

^a Outlier excluded from subsequent analyses.

* $p<0.05$.

*** $p<0.001$.

Table 4

Analyses of moderation for uncontrolled effect sizes.

Moderator	<i>n</i>	Hedges' <i>g</i>	95% CI	<i>Q(df)</i>	<i>p</i>
Study design				1.86(2)	0.39
Open trial	9	1.39 ***	1.22–1.56		
Quasi-randomized trial	2	1.81 *	0.22–3.41		
RCT	13	1.62 ***	1.32–1.92		
Population				0.003(1)	0.96
Antepartum	4	1.57 ***	0.76–2.38		
Postpartum	19	1.55 ***	1.33–1.77		
Clinician-verified diagnosis				1.82(1)	0.18
Yes	19	1.61 ***	1.40–1.82		
No	5	1.22 ***	0.69–1.75		
Measure				15.44(2)	0.000 ***
BDI	2	1.10 ***	0.94–1.27		
EPDS	16	1.62 ***	1.34–1.89		
HDRS	5	1.65 ***	1.35–1.95		
Intervention type				1.09(2)	0.58
Combination	3	1.83 ***	1.17–2.49		
Medication	4	1.46 ***	1.22–1.69		
Therapy	17	1.51 ***	1.22–1.79		
Therapy type				6.93(1)	0.008 **
Group	7	1.23 ***	0.95–1.51		
Individual	12	1.79 ***	1.48–2.10		
Therapy location				3.16(1)	0.08
Clinic	16	1.64 ***	1.36–1.92		
Home	2	2.33 ***	1.63–3.03		
CBT/IPT vs. other psychological				3.72(1)	0.05 *

Moderator	<i>n</i>	Hedges' <i>g</i>	95% CI	<i>Q(df)</i>	<i>p</i>
CBT/IPT	18	1.61 ***	1.34–1.89		
Other	2	1.11 ***	0.67–1.54		
CBT vs. other psychological				0.14(1)	0.71
CBT	7	1.63 ***	1.16–2.10		
Other	13	1.52 ***	1.21–1.83		
IPT vs. other psychological				0.23(1)	0.64
IPT	11	1.61 ***	1.25–1.97		
Other	9	1.49 ***	1.13–1.85		
CBT vs. IPT				0.003(1)	0.95
CBT	7	1.63 ***	1.16–2.10		
IPT	11	1.61 ***	1.25–1.97		

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

Table 5

Random weighted controlled effect sizes from studies comparing interventions for perinatal depression to control conditions.

Study	<i>n</i>	Hedges' <i>g</i>	SAMD
Chabrol et al. (2002)	48	2.33 ^{***}	5.38 ^a
Clark et al. (2003)	35	0.46	-0.89
Cooper et al. (2003)	184	0.39 [*]	-2.42
Grote et al. (2009)	53	1.35 ^{***}	2.14
Honey et al. (2002)	45	0.36	-1.35
Meager and Milgrom (1996)	12	0.97	0.31
Milgrom et al. (2005)	192	0.31	-2.95
Mulcahy et al. (2009)	50	0.63 [*]	-0.47
O'Hara et al. (2000)	99	1.19 ^{***}	2.13
Prendergast and Austin (2001)	37	0.31	-1.35
Spinelli and Endicott (2003)	38	0.96 ^{**}	0.57
Wickberg and Hwang (1996)	41	0.81 [*]	0.14
Wiklund et al. (2010)	33	0.51 [*]	-0.73
Yonkers et al. (2008)	31	0.60	-0.44
	<i>k</i>	Hedges' <i>g</i>	<i>Q(df)</i>
Total	14	0.76 ^{***}	40.75(13) ^{***}
Total (excluding outlier)	13	0.65 ^{***}	21.12(12) [*]
	95% CI	<i>I</i> ²	
	0.50-1.03		68.10
	0.45-0.86		43.19

^a Outlier excluded from subsequent analyses.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

Table 6

Analyses of moderation for controlled effect sizes.

Moderator	<i>n</i>	Hedges' <i>g</i>	95% CI	<i>Q(df)</i>	<i>p</i>
Control type				3.72(2)	0.16
TAU	6	0.45***	0.27–0.63		
TAU+	2	0.84	–2.04		
Wait list	3	0.93***	0.45–1.41		
Population				6.24(1)	0.01*
Antepartum	2	1.18***	0.74–1.62		
Postpartum	11	0.57***	0.38–0.75		
Clinician-verified diagnosis				0.003(1)	0.96
Yes	10	0.65***	0.40–0.90		
No	3	0.66***	0.30–1.02		
Measure				3.87(2)	0.14
BDI	2	0.35*	0.02–0.68		
EPDS	8	0.63***	0.38–0.87		
HDRS	2	0.97***	0.41–1.53		
Therapy type				3.41(2)	0.18
Combined	2	0.35*	0.02–0.68		
Group	3	0.55**	0.17–0.93		
Individual	7	0.78***	0.46–1.09		
Therapy location				2.96(1)	0.09
Clinic	9	0.73***	0.45–1.00		
Home	2	0.38*	0.09–0.67		
CBT vs. IPT				8.81(1)	0.003***
CBT	6	0.40***	0.21–0.59		
IPT	5	0.96***	0.64–1.28		

* $p < 0.05$.

.10
p<.01
***
p<.001
***