Reanalysis of Efficacy of Interpersonal Psychotherapy for Antepartum Depression Versus Parenting Education Program:
Initial Severity of Depression as a Predictor of Treatment Outcome

Margaret G. Spinelli, MD[a,*]; Jean Endicott, PhD[a]; Ray R. Goetz, PhD[a]; and Lisa S. Segre, PhD[b]

ABSTRACT

Objective: Interpersonal psychotherapy (IPT) is supported by substantial empirical evidence as a treatment for depression. Surprisingly, our recently reported randomized, single-blind, controlled clinical trial found no significant difference between interpersonal psychotherapy for antepartum depression (IPT-P) and a parenting education program (PEP) control condition for the treatment of prenatal depression. Because depression severity has been found to influence treatment response in antidepressant treatment trials, the current study reassessed IPT-P outcomes, limiting analyses to women with moderate depressive symptoms.

Method: For this reanalysis, 75 of the 110 study participants who met *DSM-IV* criteria for major depressive disorder and scored ≥ 16 on the 17-item Hamilton Depression Rating Scale (HDRS-17) from 2005 through 2011 were classified as moderately depressed. Linear mixed models were used to examine the longitudinal treatment response on the HDRS-17, the Edinburgh Postnatal Depression Scale (EPDS), and the Clinical Global Impressions Improvement (CGI-I) and Severity (CGI-S) scales.

Results: Although the longitudinal analysis did not reveal a significant interaction of treatment group and visit (ie, treatment response variation), the IPT-P group had significantly lower HDRS-17 and EPDS depression ratings than the PEP group at week 8 (respectively, $P = .008$ and $P = .046$); these scores remained low but lost significance versus those for the PEP group at week 12 due to attrition and smaller sample size. For the CGI ratings, the longitudinal analysis revealed significant interaction of treatment groups and visits for the CGI-I ($P = .021$) and CGI-S ($P = .005$) ratings. Post hoc analysis showed significant illness improvement and less illness severity for the IPT-P group as measured by the CGI ratings at weeks 8 ($P = .007$ and $P = .003$, respectively) and 12 ($P = .003$ and $P = .012$, respectively), whereas the PEP group remained relatively unchanged during the study.

Conclusions: The results of this reanalysis indicate that among women with moderate levels of depression severity, IPT-P is markedly more effective than PEP. The significance of baseline severity level in depression is important in treatment trial outcomes and considerably more important in determining treatment decisions for pregnant depressed women.

Trial Registration: ClinicalTrials.gov identifier: NCT00251043

J Clin Psychiatry 2016;77(4):535–540
dx.doi.org/10.4088/JCP.15m09787

[a]Columbia University College of Physicians and Surgeons, New York State Psychiatric Institute, New York
[b]College of Nursing, University of Iowa, Iowa City
*Corresponding author: Margaret G. Spinelli, MD, Clinical Professor of Psychiatry, Columbia University College of Physicians and Surgeons, New York State Psychiatric Institute, 1051 Riverside Dr, Box 123, New York, NY 10032 (mgs8@columbia.edu).

We recently reported[1] results of a 3-site, randomized, single-blind, controlled, parallel-design, bilingual (Spanish and English) clinical treatment trial from 2005 to 2011 to examine the comparative effectiveness of interpersonal psychotherapy (IPT) for antepartum depression (IPT-P) versus a parenting education program (PEP) control condition during 12 weeks in 102 pregnant depressed women who met *DSM-IV* criteria for major depressive disorder (MDD) and had a 17-item Hamilton Depression Rating Scale (HDRS-17) score ≥ 12. The sample was represented by equivalent numbers of Hispanic, African American, and white women. Blinded independent evaluators performed assessments. Although we hypothesized that IPT-P would be more effective than PEP for antepartum depression,[1] the results demonstrated that both IPT-P and the PEP control were equally effective for treating mild depression.

In the spirit of the National Institute of Mental Health (NIMH) Collaborative Study,[2] we have reanalyzed the data using a higher baseline severity score (≥ 16) on the HDRS-17. The NIMH Collaborative Study investigated the effectiveness of IPT, cognitive-behavioral therapy (CBT), imipramine plus clinical management, and placebo plus clinical management in the treatment of MDD. Similar to our initial primary analysis, their analysis of their total samples without regard to initial severity of illness demonstrated equal effectiveness of IPT and CBT that was not significantly less than that of imipramine.[3,4] However, secondary analysis in which patients were dichotomized on initial level of severity of depressive symptoms demonstrated that imipramine was more effective than IPT or CBT in severe depression and that psychotherapy was effective for mild-to-moderate depression. Significant differences among treatments were found for the subgroup of patients who were more severely depressed.

Randomized controlled trials (RCTs) often demonstrate a large placebo response resulting in a relatively small effect size between placebo and drug in patients suffering from MDD.[5] The failure to detect treatment effects of antidepressant treatments in placebo-controlled trials has increased over time,[6] with a mean placebo response rate increase of 7% per decade over the past 30 years.[7] Several authors have tried to identify factors and patient characteristics as predictors for increasing signal detection in clinical trials. Fournier et al[8] found that the magnitude of response of antidepressant medication compared to placebo increases with the severity of symptoms at baseline.

Data on effect sizes are primarily based on antidepressant trials, but concerns about reduced signal detection also apply

to psychotherapy trials. Head-to-head comparisons of high-quality psychotherapy studies are difficult to make. IPT trials have limitations that make meta-analytic comparisons difficult. Methodologies vary from open trials, waiting list controls, and treatment as usual in individual, couple, and group interventions. Manualization of treatments, lack of diagnostic criteria, method of assessment, and the use of independent evaluators further affect effect sizes. The effects of psychotherapy are frequently significant, but tend to be small.[9,10]

The number of psychotherapy trials for perinatal depression is too few to compare effect sizes, although Claridge[11] found positive average effect sizes in psychotherapy treatments that addressed perinatal depression when treatments were compared with control groups. Effect sizes tend to be larger among individual IPT studies and clinical samples.

In the first study to provide evidence-based guidelines for interpreting HDRS-17 scores both cross-sectionally for severity and longitudinally for change in severity, Furukawa et al[12] found that scores between 8 and 15 signify "mildly ill" and those between 16 and 26 represent "moderately ill" patients. With these guidelines in mind, we explored the possibility that the PEP, a didactic control condition of individual therapist-led 45-minute weekly educational sessions[13] was just as effective as IPT-P in the primary analysis because depressive symptoms were mild and close to the cutoff score for admission. We hypothesized that baseline disease severity of patients was the most important factor for the finding that PEP and IPT-P were of equal benefit. We postulated that IPT-P would be more efficacious than PEP in women with HDRS-17 baseline scores ≥ 16, representing moderate depression.[12]

We also considered the possibility that somatic symptoms artificially increased HDRS scores. Pereira et al[14] demonstrated that pregnant women experience significant somatic changes even if not clinically depressed. Subjects with milder depression may have had elevated HDRS scores because of the difficulty in distinguishing the normal physiologic discomforts that characterize the gestational period from the somatic symptoms associated with a depressive episode, and consequently met the baseline HDRS-17 cutoff score for inclusion.

While the literature on failure to detect large effect sizes has focused on medication trials, we are extrapolating these findings for the purposes of our psychotherapy study trial design to determine the potential impact of baseline depression severity on outcome. We have reanalyzed the data using an initial HDRS-17 score ≥ 16 as an inclusion criterion. Because trial designs have significant impacts on outcome, we hypothesized that IPT-P would be more effective than PEP in women who had more severe depression at entry.

METHOD

Study Design

More than 479 prospective research participants were referred to the Maternal Mental Health Program at the New York State Psychiatric Institute from the obstetrics departments of 3 metropolitan hospitals in New York City, each representing various racial, ethnic, and socioeconomic groups.[1]

Women who met *DSM-IV* criteria for a major depressive episode using the Structured Clinical Interview for *DSM-IV*, Research Version, Patient Edition (SCID-I/P),[15] and had a minimum score of 12 on the HDRS-17[16] were invited into the treatment phase of the study (ClinicalTrials.gov identifier: NCT00251043). One hundred forty-two women met initial study criteria. Seventy-five met the criterion for moderate depression (HDRS-17 score ≥ 16) and were included in this reanalysis. A more comprehensive discussion of study design and therapist training has been previously published.[1]

Prior to randomization, patients received a complete psychiatric evaluation and were assessed with the SCID-I/P, the Edinburgh Postnatal Depression Scale (EPDS),[17] the HDRS-17, and the Clinical Global Impressions scale (CGI).[18] Randomization was stratified by site, ethnicity, and trimester of pregnancy using a block size of 4. Blinded independent raters administered the HDRS-17, EPDS, and CGI at baseline (in person) and weeks 4, 8, and 12 (in person or by telephone). English- and Spanish-speaking women between 12 and 33 weeks' gestation and 18–45 years of age were included in the study. All subjects signed informed consent in English or Spanish. The Institutional Review Boards at each institution approved the study.

Statistical Analyses

The subjects of this reanalysis are the individuals from each treatment group (IPT-P and PEP) whose baseline HDRS-17 ratings were ≥ 16. The demographic measures were examined across the 2 treatment groups using the *t* test statistic for the continuous measures and the χ^2 statistic and the Fisher exact test for the categorical measures. To examine the longitudinal treatment response data (HDRS-17, EPDS, and CGI Severity [CGI-S] and Improvement [CGI-I] ratings), we used linear mixed models (LMM) procedures available in IBM SPSS Statistics (Release 21). For each LMM procedure, the independent variable was treatment group membership, and age, race, education level, and income were included as covariates. The primary outcome was score on

Clinical Points

- The decision to treat depression during pregnancy involves a risk/benefit analysis and tends to be a difficult one for the clinician and the patient.
- The gravity of depression symptoms is a good indicator for the decision to use a non-pharmacologic method of treatment such as psychotherapy rather than antidepressant medication. Current guidelines from the American Psychiatric Association and the American College of Obstetricians and Gynecologists recommend the use of psychotherapy for mild-to-moderate depression in pregnancy.
- Interpersonal psychotherapy demonstrates efficacy for the treatment of mild-to-moderate depressive symptoms in the antepartum period.

the HDRS-17. The main focus of these analyses was the interaction term of treatment group and time, which would indicate treatment group differences over the course of treatment. A 2-tailed α of .05 was used for each statistical test.

RESULTS

Sample

In the original report, which included 110 cases (IPT-P N = 63 and PEP N = 47), all statistical effects indicative of treatment group–by-visit differences, for each of the outcome measures, were nonsignificant. For the current reexamination, we selected the individuals from each group who presented with baseline HDRS-17 item ratings ≥ 16 (IPT-P, N = 46; PEP, N = 29) as obtained by independent evaluators.

Demographics and other participant characteristics are presented and compared across the treatment groups in Table 1. There were no significant treatment group differences in age or gestation. Education level, race, marital and immigration status, and income were well dispersed within and across the 2 treatment groups, showing no significant distribution differences. There was a trend for differences between the 2 treatment groups for previous pregnancies and previous depression: the PEP cases exhibited more previous episodes of depression (Fisher exact test $P = .049$) and a trend for more previous pregnancies (χ^2 $P = .058$).

Table 1. Demographic Measures of the Participants in the Depression During Pregnancy Study Who Presented With Baseline HDRS-17 Scores ≥ 16[a]

	IPT-P Treated (N = 46)		PEP Treated (N = 29)		Group Effect		
Demographic Measures	Mean	SD	Mean	SD	*t*	*df*	*P*
Age, y	30.9	6.6	29.7	6.6	0.78	73	.440
Gestation, wk	21.8	6.2	23.1	7.0	0.85	72	.400
Baseline HDRS-17 score	19.2	2.7	20.3	3.4	1.46	73	.148
Categorical Measures	%	n	%	n	χ^2	*df*	*P*
Education level					3.43	4	.480
Less than high school	20	9	24	7			
High school/GED	13	6	14	4			
Some college	22	10	35	10			
BA/BS degree	26	12	21	6			
Graduate school	20	9	7	2			
Race/ethnicity					2.39	2	.303
Hispanic	28	13	45	13			
Black	30	14	28	8			
White	41	19	28	8			
Marital status					0.26	1	.608
Married or common law	54	25	48	14			
Single/divorced/separated	46	21	52	15			
Income, US $					3.01	4	.557
< 15,000	24	11	21	6			
15,000–24,999	7	3	17	5			
25,000–39,999	24	11	28	8			
40,000–59,999	9	4	3	1			
≥ 60,000	37	17	31	9			
Immigration status					0.42	1	.515
US born	65	30	72	21			
Immigrant	35	16	28	8			
No. of previous pregnancies					5.71	2	.058
0	39	18	14	4			
1 or 2	46	21	69	20			
≥ 3	15	7	17	5			
Previous episodes of depression*					5.29	1	.071
No	96	44	79	23			
Yes	4	2	21	6			

[a]Test statistics are the *t* test for continuous measures and the χ^2 test for categorical measures.
*Fisher exact test $P = .049$.
Abbreviations: GED = General Educational Development, HDRS-17 = 17-item Hamilton Depression Rating Scale, IPT-P = interpersonal psychotherapy for antepartum depression, PEP = parenting education program.

General Linear Mixed Models Analyses and Post Hoc Univariate Analyses of Covariance

For the HDRS-17 total score, there was a significant group effect ($F_{1,157} = 7.47$, $P = .007$) as well as a significant across-visits effect ($F_{3,91} = 120.72$, $P < .001$). Although the interaction term for group by visit was not significant ($F_{3,91} = 1.42$, $P = .243$), it is clear from Figure 1 that the IPT-P group appears to have lower HDRS-17 depression ratings at weeks 8 and 12 than the PEP group. Post hoc univariate analyses of covariance at weeks 8 and 12 revealed significantly lower depression levels for the IPT-P group at week 8 ($F_{1,54} = 7.47$, $P = .008$), but not at week 12 ($F_{1,39} = 2.47$, $P = .124$).

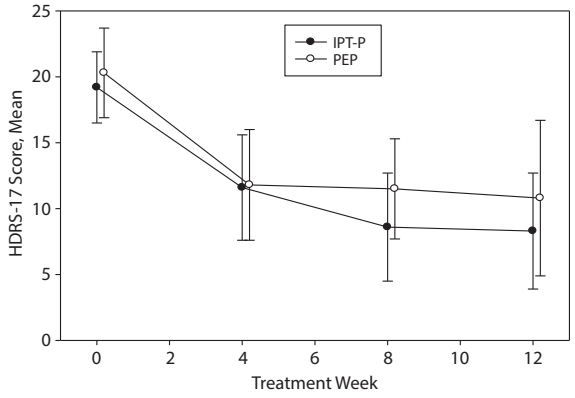
The results for the EPDS (see Figure 2) were similar. There was a significant group effect ($F_{1,172} = 7.58$, $P = .007$) as well as a significant across-visits effect ($F_{3,83} = 59.00$, $P < .001$); however, the interaction term for group by visit was not significant ($F_{3,83} = 0.29$, $P = .835$). Post hoc analyses showed that the IPT-P group exhibited significantly lower EPDS scores at baseline ($F_{1,69} = 5.93$, $P = .018$) and at week 8 ($F_{1,49} = 4.19$, $P = .046$), but not at weeks 4 and 12.

CGI-I ratings were obtained at weeks 4, 8, and 12 (Figure 3). The generalized LMM (GLMM) revealed a significant group effect ($F_{1,132} = 11.41$, $P = .001$), a significant across-visits effect ($F_{2,84} = 3.11$, $P = .050$), and a significant group-by-visit interaction ($F_{2,84} = 4.03$, $P = .021$). It seems apparent from Figure 3 that the IPT-P group shows a clear improvement trajectory while the PEP group remains unchanged. Post hoc analyses showed significant differences between the 2 treatment groups at week 8 ($F_{1,50} = 7.87$, $P = .007$) and week 12 ($F_{1,39} = 10.15$, $P = .003$) and no difference at week 4. For the CGI-S ratings (Figure 4), the GLMM analysis revealed a significant group effect ($F_{1,150} = 13.26$, $P < .001$), a significant across-visits effect ($F_{3,93} = 25.74$, $P < .001$), and a significant group-by-visit interaction ($F_{3,94} = 4.63$, $P = .005$). The post hoc analyses revealed significantly lower CGI-S scores for the IPT-P group at weeks 8 ($F_{1,50} = 9.44$, $P = .003$) and 12 ($F_{1,39} = 7.01$, $P = .012$), with no differences at baseline and week 4. The PEP group CGI-S score remained above 3.00 throughout the course of treatment, showing minimal decline in severity of illness.

DISCUSSION

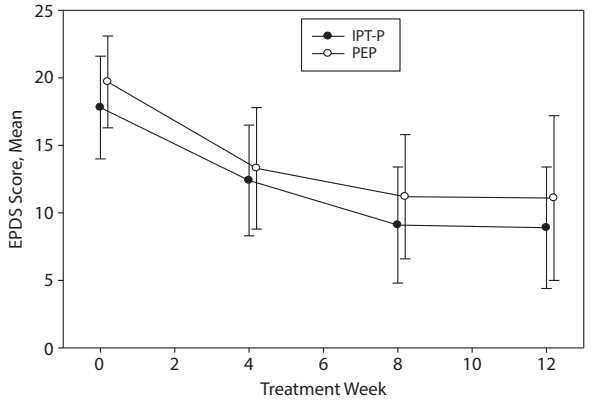
In recent years, discussion has focused on the design features of current MDD studies that might artificially inflate placebo

Figure 1. Comparison of 17-Item Hamilton Depression Rating Scale (HDRS-17) Scores of Pregnant Women During Treatment With Interpersonal Psychotherapy for Antepartum Depression (IPT-P) or Parenting Education Program (PEP) at Baseline and Weeks 4, 8, and 12 of Treatment[a,b]



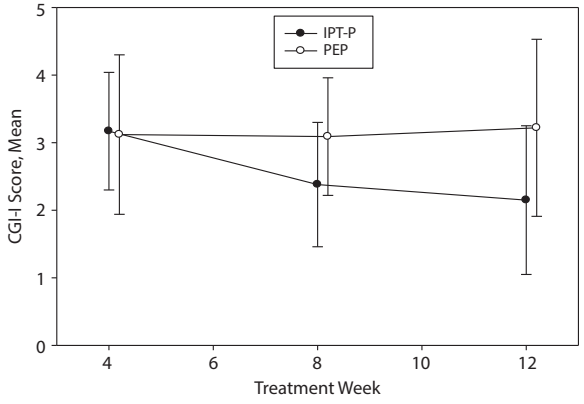
[a]Ratings were performed by independent evaluators. Cases were selected based on HDRS-17 score ≥ 16 at baseline.
[b]Error bars indicate standard deviation (SD).

Figure 2. Comparison of Edinburgh Postnatal Depression Scale (EPDS) Scores of Pregnant Women During Treatment With Interpersonal Psychotherapy for Antepartum Depression (IPT-P) or Parenting Education Program (PEP) at Baseline and Weeks 4, 8, and 12 of Treatment[a,b]



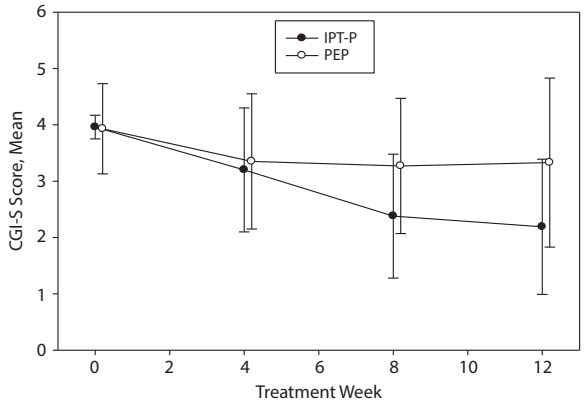
[a]Ratings were performed by independent evaluators. Cases were selected based on 17-item Hamilton Depression Rating Scale score ≥ 16 at baseline.
[b]Error bars indicate standard deviation (SD).

Figure 3. Comparison of Clinical Global Impressions-Improvement Scale (CGI-I) Scores of Pregnant Women During Treatment With Interpersonal Psychotherapy for Antepartum Depression (IPT-P) or Parenting Education Program (PEP) at Weeks 4, 8, and 12 of Treatment[a,b]



[a]Ratings were performed by independent evaluators. Cases were selected based on 17-item Hamilton Depression Rating Scale score ≥ 16 at baseline.
[b]Error bars indicate standard deviation (SD).

Figure 4. Comparison of Clinical Global Impressions-Severity (CGI-S) Scores of Pregnant Women During Treatment With Interpersonal Psychotherapy for Antepartum Depression (IPT-P) or Parenting Education Program (PEP) at Baseline and Weeks 4, 8, and 12 of Treatment[a,b]



[a]Ratings were performed by independent evaluators. Cases were selected based on 17-item Hamilton Depression Rating Scale score ≥ 16 at baseline.
[b]Error bars indicate standard deviation (SD).

responses. Methodological factors such as patient selection contribute to the increasing rates of failed antidepressant trials.[3] Our initial patient sample in our psychotherapy study included women with a wide range of baseline depression severity, including mild depression. In this follow-up second analysis, we dichotomized the patient sample according to severity criteria. In contrast to the less severely depressed subgroup, there was evidence of IPT-P efficacy for moderate depression. This finding supports the influential role that severity of depression plays in clinical trial outcomes as well as the clinical discussions and decisions about treatment options during pregnancy.

There were significant group differences across the clinical treatment trial. The IPT-P group had lower mean HDRS-17 and EPDS ratings than the PEP group at weeks 8 and 12. Post hoc analysis demonstrated that IPT-P was significantly more effective than PEP at week 8 (effect size = 0.74), but lost significance at week 12 (effect size = 0.50) possibly due to a smaller effect size and/or attrition/smaller sample size. The attrition rates for the IPT-P group from weeks 0–4, 5–8, and 9–12 were respectively 8.7%, 10.9%, and 21.7%, while attrition rates for the PEP group were respectively 0%, 20.7%, and 21.7%. In the weeks 8–12, the IPT-P group lost twice the number of subjects compared to weeks 0–4 and 4–8 (Table 2).

Table 2. Visit-to-Visit Attrition Rates From Baseline

Week of Visit	IPT-P (baseline N = 46) n (attrition %[a])	PEP (baseline N = 29) n (attrition %[a])
0–4	42 (8.7)	29 (0)
5–8	37 (10.9)	23 (20.7)
9–12	27 (21.7)	18 (17.2)
	Attrition, n (%[a])	Attrition, n (%[a])
Total	19 (41.3)	11 (37.9)

[a]Percentages are determined using the baseline N as the denominator.
Abbreviations: IPT-P = interpersonal psychotherapy for antepartum depression, PEP = parenting education program.

In contrast to our initial analysis, this reanalysis determined that the IPT-P group had significantly greater global improvement in illness throughout the study and at the week 12 endpoint compared to the PEP group. The PEP control group failed to demonstrate any improvement over the course of the study. For the CGI-S ratings, the IPT-P group had significantly less severe illness over the course of treatment and week 12 while the PEP group showed almost no decline in illness severity.

The global quality of the CGI may explain the significant differences between groups, which, unlike differences on the HDRS-17 and the EPDS, were sustained through the end of treatment week 12. While the HDRS-17 and EPDS measure the presence and severity of discrete illness variables, eg, depression, global ratings of illness provide valuable information about the overall clinical picture of each subject when assessing the efficacy of a treatment by providing more comprehensive information about patients' clinical status.[19] Unlike scores on the HDRS-17 and the EPDS, the CGI-S and CGI-I ratings are not merely a composite of symptom ratings of depression; rather, they incorporate core symptomatology of the disorder. The CGI indicates discomfort, distress, and impairment[19] and reflects cues from the patient regarding sense of well-being and functional capacity. The CGI is considered by many to be the gold standard for designating responder status. These findings replicate the initial pilot controlled trial that demonstrated the efficacy of IPT-P compared to PEP for antepartum depression.[13]

Severity scores have important treatment and research implications since mild depression requires a less intensive treatment such as a support group, while moderate and severe depression may require psychotherapy, medication, or both.[20] These treatment decisions are particularly critical during pregnancy to determine the most effective treatment with the least number of adverse effects for mother and fetus.

The high number of patients with mild depression included in clinical trials can drag down the overall effect size of the study and call into question the influence of depression severity at baseline; this depression severity may be due to methodological artifacts of randomized controlled trials, rater bias at the time of study entry, or artificially elevated scores in the case of pregnancy.

Mancini et al[5] found a decrease in signal detection with patients who had baseline HDRS scores that were closest to the cutoff scores for enrollment. In our original analysis, we chose a baseline cutoff score of HDRS-17 ≥ 12 based on the current treatment guidelines for perinatal depression, which recommend psychotherapy for mild-to-moderate depression.[21] We chose the PEP control group in lieu of a waiting list control to determine the specific effects of IPT-P. In the initial analysis, psychoeducation about pregnancy, fetal growth, and delivery provided a select treatment effect for pregnant women with mild depression. Other educational models have established benefit for treatment[22] and prevention[23] of perinatal depression. In addition, the quality of the therapeutic relationship accounts for about 30% of the variance in outcome.[24] These factors very likely enriched the therapeutic effect of PEP, further contributing to loss of significance in our original analysis.

In addition to the strength of the control condition, other factors may account for the strong placebo response. Mundt et al[25] opined that inclusion of participants with mild depression might be due to the tendency to unconsciously inflate baseline scores to facilitate enrollment. Such enthusiasm and unconscious bias allow for a preponderance of baseline cutoff scores.

Another factor that could affect depression scores is artificial inflation due to endorsement of somatic symptoms of pregnancy, which inflate HDRS scores enough to meet criteria for entry into the study. The EPDS does not have a somatic component. Both EPDS and HDRS-17 have excellent diagnostic validity and similar sensitivities and specificities and are highly predictive of MDD when used in the pregnant or postpartum populations.[26,27] Since there are no observer-rated scales that do not include somatic symptoms, the HDRS-17 is an important complement to self-report measures such as the EPDS. The question of whether somatic symptoms are valid indicators of depression in pregnant women suggests exclusion when assessing pregnant women for depression.[28] On the other hand, some contend that the alternative may lead to a loss of vital information.[29] Ji et al[27] suggested that the presence of these factors did not elevate their scores and that women tend to incorporate their own opinions about the etiology of symptoms. In contrast, Ross et al[30] demonstrated that specific somatic items on the HDRS do interfere with the scale's potential for measuring prenatal depression, suggesting that rating scales with somatic symptoms should be used with caution when assessing antepartum depression. To offset this problem in the HDRS-17, we asked subjects whether a particular symptom was attributed to pregnancy (not scored) or to depression (endorse symptom). In addition, post hoc analysis revealed no differences between HDRS scores between groups when somatic symptoms were excluded.

A limitation of our study is the use of certified social workers for both IPT-P and PEP due to the limited numbers of bilingual practitioners available; however, our videotape reviews support adherence to the protocol.

CONCLUSION

The central influence of our study lies in the designation of treatment guidelines for depression in pregnancy. Our

reanalysis replicates our initial randomized clinical trial of IPT-P efficacy for antenatal depression and endorses treatment guidelines of the American Psychiatric Association and the American College of Obstetricians and Gynecologists[21] that recommend IPT-P for mild-to-moderate depression during pregnancy. Since many women prefer to avoid fetal antidepressant exposure during pregnancy, our findings are critical and emphasize the identification of baseline depression severity evaluation by the clinician to provide informed consent and make appropriate recommendations for treatment. In summary, we have replicated the efficacy of IPT-P for moderate depression during pregnancy and confirm the recommendation for use in antepartum depression.

In addition, we have demonstrated that pretreatment severity of illness is a variable that has a significant influence on treatment outcome. Baseline level of depression is a differential predictor of response at termination of treatment[3,31] and a likely cause of large placebo effects in both medication and psychotherapy trials. The efficacy of either treatment, compared with placebo, often increases with the severity of depressive symptoms, being more substantial in patients with more severe depression. We suggest that restricting primary analysis of clinical trials to a population with a higher cutoff score representing moderate depression will lead to better signal detection of psychotherapy efficacy. Rating scales that include somatic symptoms should be used with caution when assessing antepartum depression.

Submitted: January 12, 2015; accepted June 30, 2015.

Drug names: imipramine (Tofranil and others).

Potential conflicts of interest: **Dr Spinelli** has received research support from NIMH and has been a consultant for Pfizer and Forest. **Dr Endicott** has received research support from Cyberonics, has served as a speaker or a member of the advisory board for AstraZeneca, Bayer, Amgen, and Pfizer; and is a stock shareholder for Cyberonics. **Drs Goetz** and **Segre** report no conflicts of interest.

Funding/support: This study was supported by a research grant from the National Institute of Mental Health.

Role of the sponsor: No active role.

REFERENCES

1. Spinelli MG, Endicott J, Leon AC, et al. A controlled clinical treatment trial of interpersonal psychotherapy for depressed pregnant women at 3 New York City sites. *J Clin Psychiatry*. 2013;74(4):393–399.
2. Elkin I, Shea MT, Watkins JT, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program. general effectiveness of treatments. *Arch Gen Psychiatry*. 1989;46(11):971–983.
3. Elkin I, Gibbons RD, Shea MT, et al. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol*. 1995;63(5):841–847.
4. Klein DF, Ross DC. Reanalysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program General Effectiveness Report. *Neuropsychopharmacology*. 1993;8(3):241–251.
5. Mancini M, Wade AG, Perugi G, et al. Impact of patient selection and study characteristics on signal detection in placebo-controlled trials with antidepressants. *J Psychiatr Res*. 2014;51:21–29.
6. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry*. 2013;170(7):723–733.
7. Walsh BT, Seidman SN, Sysko R, et al. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA*. 2002;287(14):1840–1847.
8. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010;303(1):47–53.
9. Cuijpers P, van Straten A, Bohlmeijer E, et al. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol Med*. 2010;40(2):211–223.
10. van Hees MLJM, Rotter T, Ellermann T, et al. The effectiveness of individual interpersonal psychotherapy as a treatment for major depressive disorder in adult outpatients: a systematic review. *BMC Psychiatry*. 2013;13(1):22.
11. Claridge AM. Efficacy of systemically oriented psychotherapies in the treatment of perinatal depression: a meta-analysis. *Arch Women Ment Health*. 2014;17(1):3–15.
12. Furukawa TA, Akechi T, Azuma H, et al. Evidence-based guidelines for interpretation of the Hamilton Rating Scale for Depression. *J Clin Psychopharmacol*. 2007;27(5):531–534.
13. Spinelli MG, Endicott J. Controlled clinical trial of interpersonal psychotherapy versus parenting education program for depressed pregnant women. *Am J Psychiatry*. 2003;160(3):555–562.
14. Pereira AT, Marques M, Soares MJ, et al. Profile of depressive symptoms in women in the perinatal and outside the perinatal period: similar or not? *J Affect Disord*. 2014;166:71–78.
15. First MB, Spitzer RL, Gibbon M, et al. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P)*. New York, NY: Biometrics Research, New York State Psychiatric Institute; 2002.
16. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23(1):56–62.
17. Klein MH, Essex MJ. Pregnant or depressed? the effect of overlap between symptoms of depression and somatic complaints of pregnancy on rates of major depression in the second trimester. *Depression*. 1994;2(6):308–314.
18. Petkova E, Quitkin FM, McGrath PJ, et al. A method to quantify rater bias in antidepressant trials. *Neuropsychopharmacology*. 2000;22(6):559–565.
19. Leon AC, Shear MK, Klerman GL, et al. A comparison of symptom determinants of patient and clinician global ratings in patients with panic disorder and depression. *J Clin Psychopharmacol*. 1993;13(5):327–331.
20. Pollack LO, Segre LS, O'Hara MW, et al. Postpartum depression: the distribution of severity in a community sample. *Arch Women Ment Health*. 2011;14(4):363–364.
21. Yonkers KA, Wisner KL, Stewart DE, et al. The management of depression during pregnancy: a report from the American Psychiatric Association and the American College of Obstetricians and Gynecologists. *Gen Hosp Psychiatry*. 2009;31(5):403–413.
22. Howell EA, Balbierz A, Wang J, et al. Reducing postpartum depressive symptoms among black and Latina mothers: a randomized controlled trial. *Obstet Gynecol*. 2012;119(5):942–949.
23. Lara MA, Navarro C, Navarrete L. Outcome results of a psycho-educational intervention in pregnancy to prevent PPD: a randomized control trial. *J Affect Disord*. 2010;122(1–2):109–117.
24. Assay TP, Lambert MJ. The empirical case for the common factors in therapy: quantitative findings. In: Hubbell MA, Duncan BL, Miller SD, eds. *The Heart and Soul of Change. What Works in Therapy?* Washington, DC: American Psychological Association; 1998:23–55.
25. Mundt JC, Greist JH, Jefferson JW, et al. Is it easier to find what you are looking for if you think you know what it looks like? *J Clin Psychopharmacol*. 2007;27(2):121–125.
26. Thompson WM, Harris B, Lazarus J, et al. A comparison of the performance of rating scales used in the diagnosis of postnatal depression. *Acta Psychiatr Scand*. 1998;98(3):224–227.
27. Ji S, Long Q, Newport DJ, et al. Validity of depression rating scales during pregnancy and the postpartum period: impact of trimester and parity. *J Psychiatr Res*. 2011;45(2):213–219.
28. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry*. 1987;150(6):782–786.
29. Bernstein IH, Rush AJ, Yonkers K, et al. Symptom features of postpartum depression: are they distinct? *Depress Anxiety*. 2008;25(1):20–26.
30. Ross LE, Gilbert Evans SE, Sellers EM, et al. Measurement issues in postpartum depression part 2: assessment of somatic symptoms using the Hamilton Rating Scale for Depression. *Arch Women Ment Health*. 2003;6(1):59–64.
31. Marchesi C, De Panfilis C, Tonna M, et al. Is placebo useful in the treatment of major depression in clinical practice? *Neuropsychiatr Dis Treat*. 2013;9:915–920.

Editor's Note: We encourage authors to submit papers for consideration as a part of our Focus on Women's Mental Health section. Please contact Marlene P. Freeman, MD, at mfreeman@psychiatrist.com.