Published in final edited form as:

Arch Womens Ment Health. 2014 April ; 17(2): 97–105. doi:10.1007/s00737-013-0385-4.

Representativeness of obstetric patients who participate in perinatal depression research: findings from the Women's Mental Health and Infants Program (WMHIP) integrated dataset

Danielle M. Novick,

Department of Psychiatry, University of Michigan, 4250 Plymouth Road, SPC 5769, Ann Arbor, MI 48109, USA

University Drive C, VA Pittsburgh Healthcare System, Pittsburgh, PA 15240, USA

Lucy Allbaugh,

Department of Psychology, Miami University, Psychology Building, 100, Oxford, OH 45056, USA

Zhuo Zhao,

Department of Psychiatry, University of Michigan, 4250 Plymouth Road, SPC 5769, Ann Arbor, MI 48109, USA

Erin Henshaw,

Denison University, 100 West College Street, Granville, OH 43023, USA

Delia M. Vazquez,

Department of Pediatrics–Endocrinology, University of Michigan, 1500 E. Medical Center Drive, SPC1871, Ann Arbor, MI 48109, USA

Roseanne Armitage,

Department of Psychiatry, University of Michigan, 4250 Plymouth Road, SPC 5769, Ann Arbor, MI 48109, USA

Heather Flynn,

Department of Medical Humanities and Social Sciences, Florida State University, 1115 West Call Street, Tallahassee, FL 32306-4300, USA

Department of Psychiatry, University of Michigan, 4250 Plymouth Road, SPC 5769, Ann Arbor, MI 48109, USA

Women's Mental Health and Infants Program (WMHIP)

Department of Psychiatry, University of Michigan, 4250 Plymouth Road, SPC 5769, Ann Arbor, MI 48109, USA

Abstract

The aims of this study were to evaluate the feasibility of integrating archival datasets from depression projects involving pregnant women recruited from obstetric clinics and then assess the representativeness of the integrated dataset. Datasets from six studies were standardized and

dnovick@med.umich.edu.

Conflict of interest The authors declare that they have no conflicts of interest and no competing financial interests.

integrated. Chi-square, t -, and Wilcoxon rank-sum tests were used to compare characteristics between women who completed a depression screening questionnaire (DSQ) and were (1) eligible and ineligible for research participation and (2) eligible women who accepted and declined participation. The integrated dataset comprises 9,112 pregnant women, of whom 71.0 % ($n=6,472$) were ineligible for participation because their DSQ scores indicated no-to-minimal depressive symptoms (NDS). Among the 23.9 % (2,176) of women identified as eligible, in part, because their DSQ scores indicated elevated levels of depressive symptoms (EDS), 29.6 % (644) of women participated (P-EDS) and 47.6 % (1,036) of women did not participate (D-EDS). While the NDS and EDS groups were significantly different on almost all variables, the P-EDS and D-EDS groups were significantly different on only a few variables. Compared to the D-EDS group, the P-EDS group was earlier in pregnancy and, on the Edinburgh Postnatal Depression Screen, was more likely to endorse impaired “ability to laugh” and “enjoy oneself”, and endorse at greater severity “ability to laugh.” It is a reasonable and feasible strategy to integrate thematically similar datasets to increase statistical power. Additionally, typical recruitment strategies for minimal risk perinatal depression research at obstetric clinics, during routine prenatal care visits, appear to produce an externally valid study cohort.

Keywords

Perinatal; Depression; Selection bias; Validity; Recruitment

Introduction

The point prevalence of prenatal and postnatal major and minor depression has been estimated to be as high as 11.0 and 12.9 %, respectively (Gaynes et al. 2005), and is associated with great disease burden (Yonkers et al., 2011). When depression emerges or recurs during or following pregnancy, the consequences may have short- and long-term psychological and biological detriment to not only the woman but also her developing fetus or infant (National Research Council and Institute of Medicine NRC 2009). Accordingly, a number of guidelines and reports include a recommendation that health professionals identify and intervene with high-risk populations (US Department of Health and Human Services Agency US DHHS for Healthcare Research and Quality UDHHS AHRQ 2011). Despite growing awareness and research efforts, the prevalence of prenatal and postnatal depression is unchanged, and they remain significant and costly public health problems (National Research Council and Institute of Medicine NRC 2009). Therefore, it is critical for the perinatal depression field to prioritize strategies that maximize the public health impact of its research.

According to the National Advisory Mental Health Council’s Workgroup (2010), one strategy to advance the field is for researchers to “work together to build and share resources such as standardized assessments, large cohorts, and shared databases.” The workgroup emphasizes the need for data standardization, integration, and sharing to drastically advance scientific discoveries and increase the public health impact of research. Against this backdrop, to facilitate interdisciplinary collaboration and data sharing in perinatal research,

the Women's Mental Health and Infants Program (WMHIP) was established at the University of Michigan Depression Center.

Pooling data from multiple projects has the potential to substantially increase sample size, thereby enhancing statistical power to address high impact research questions that individual projects may not be able to investigate (National Advisory Mental Health Council's Workgroup NAMHC 2010). One such question is whether selection bias and similar threats to external validity limit generalizability, and therefore, the clinical utility of research findings. In the case of perinatal depression research, obstetric clinics are a common setting for recruitment of pregnant and postpartum women into research projects, and depression screening questionnaires are a common tool for determining women's potential eligibility for research participation (Agency for Healthcare Research and Quality AHRQ 2012). One supposition of this recruitment method is that, since all women who have the characteristics of interest have an equal chance of being selected, samples will be representative of the defined populations. Yet, samples comprise women who are not only eligible but also agree to participate in research, whereas populations comprise women who do and do not meet eligibility criteria and do and do not agree to participate in research. The sociodemographic and clinical characteristics of these groups may be vastly different due to selection bias. Since selection bias may threaten external validity, and consequently, limit the public health impact of research, it would be useful to the perinatal depression field to investigate it.

Selection bias is a major concern in general depression research—especially research involving treatment—because external validity is often sacrificed in favor of internal validity (Robinson and Rickels 2000; Stirman et al. 2003; Zimmerman et al. 2004). Indeed, investigators have demonstrated that symptomatic research participants who meet strict inclusion/ exclusion criteria are rarely representative of the typical patient health professionals evaluate and treat in their clinical practice (Brody et al. 2011; Preskorn 2003). Thus, insights obtained from rigorous, empirical investigations may have limited public health impact or clinical utility (Brody et al. 2011; Preskorn 2003; Stirman et al. 2003; Zimmerman et al. 2004).

In this report, we detail WMHIP's foundational efforts to standardize and integrate archival datasets from six separate but thematically related depression projects involving pregnant women and "standard" recruitment and screening methods. Recruitment and screening procedures certainly vary within the perinatal depression field. Yet, there is an increasingly common obstetric-based approach employed by many investigators (Agency for Healthcare Research and Quality AHRQ 2012), and the projects selected for this report follow procedures similar to these national efforts. We define this standard recruitment and screening approach as including (1) recruitment that primarily occurs in obstetric clinics during women's routine prenatal care appointments, (2) initial screening that involves a depression screening questionnaire (DSQ), and (3) research eligibility criteria that requires meeting or exceeding a specified DSQ cutoff score. With the new WMHIP integrated dataset and substantially increased sample size, we were interested in examining selection bias. Selection bias is defined here as any systematic difference in sociodemographic or clinical characteristics between pregnant women who meet research eligibility criteria and do or do not participate in the research project.

Comparisons between women who completed a DSQ and were eligible and ineligible for research participation will allow us to describe our sample and assess whether the integrated sample of women is representative of the targeted population. Comparisons between eligible women who did and did not participate will allow us to investigate selection bias. If the study cohort is externally valid, then (1) the percentage of pregnant women with DSQ scores indicating elevated levels of depressive symptoms should reflect established prevalence rates, (2) pregnant women with and without elevated levels of depressive symptoms should differ significantly on a range of sociodemographic and clinical variables historically associated with depression risk, and (3) pregnant women with elevated levels of depressive symptoms who do or do not participate in a research project should be similar on all variables, including type or severity of depressive symptoms.

Materials and methods

This study was a secondary analysis of archival datasets from six studies (NIMH MH063880, NIMH MH065062, NIHG M01 RR00042). For each study, the University of Michigan's Biomedical Institutional Review Board approved all recruitment, assessment, and research procedures. All participants provided informed consent. No data in the integrated dataset was linkable to the source study data, and therefore, all data was de-identified and anonymous.

Participants

The integrated dataset includes information from 9,112 pregnant women who completed a DSQ between 1999 and 2011. For all studies, targeted in-person recruitment was used within obstetric clinics within both a University-affiliated health system, as well as an urban health system that primarily serves individuals receiving Medicaid. In all settings, the DSQ was part of women's routine clinical care. Whereas most women were approached by research staff members about research participation during a healthcare appointment, some women contacted the research team after seeing advertised materials in the clinics.

Research staff members identified all women as potentially eligible or ineligible for research participation in one of six depression projects based on project-specific eligibility criteria that included elevated DSQ scores. These six studies are referred to as "source studies." Table 1 provides an overview of each source study and full eligibility criteria. Women with elevated DSQ scores were invited to participate in full screening procedures for a specified research project. The full screening procedures varied in process and complexity across studies, and some studies had no screening procedures beyond the DSQ and ability to provide informed consent. Most of the women who did not meet initial screening criteria (i.e., elevated DSQ scores) or who met criteria but declined to participate in further screening provided consent for their deidentified information to be added to the research database. Thus, each archival dataset included at least basic sociodemographic and clinical information about the population. Archival datasets did not include the number of women who requested that their de-identified information not be added to the research database; this number, however, was remembered as very low (personal communication with PIs).

Inclusion and exclusion criteria for the integrated dataset

To be included in the integrated dataset, women needed to (1) be pregnant and 18 years of age or older at the time of source study screening and (2) have completed at least one of two DSQs: Edinburgh Postnatal Depression Screen (EPDS) (Cox 1983) or Center for Epidemiologic Studies Depression Scale (CES-D) (Radloff 1977).

Participant classification

Women were classified into groups according to their DSQ scores and final source study participation status. Women with DSQ scores indicating “elevated levels of depressive symptoms” (EDS) were identified as eligible, whereas women with DSQ scores indicating “no-to-minimal depressive symptoms” (NDS) were identified as ineligible. EDS was defined by the source study: four source studies required a score of 10 or above on the EPDS, one source study a score of 12 or above on the EPDS, and one source study a score of 16 or above on the CES-D. In general, these cut points were adopted by each source study to maximize sensitivity (Agency for Healthcare Research and Quality AHRQ 2012).

Women in the EDS group who completed all screening procedures, provided written informed consent, and completed at least one study procedure post-screening were categorized as “Participants with EDS” (P-EDS). Women in the EDS group who did not meet these criteria were classified as “Did not participate with EDS” (D-EDS). To clarify, women in the D-EDS group completed a DSQ as part of their routine clinical care but did not begin or enroll in a source study.

Procedure

The dataset integration evolved through four phases. During phase 1, all source study investigators met to establish data sharing agreements, review consent and human use issues, and determine procedures. During phase 2, a subset of investigators compiled all source datasets and codebooks. Then, each variable was classified in the source datasets as belonging to specific categories. For instance, all variables related to income were labeled “Household Income.” All variables related to history of depression were labeled “Illness History.” Key categories are listed in Tables 2 and 3 under “Variable.” This classification allowed us to identify overlapping or similar variables across source datasets. During phase 3, when possible, levels of the overlapping or similar variables were standardized and recoded. For example, three source studies may have coded household income as (a) under \$14,999/year, (b) between \$15,000 and \$19,999/year, etc., whereas two source studies may have coded household income as (c) under \$10,000/year, (d) between \$10,000 and \$20,000/year, etc., and one source study may have code household income as (e) under \$800/month and (f) between \$800 and \$1,667/ month., etc. To create one “Household Income” variable for the integrated dataset, levels were collapsed and/or recalculated to be standardized, and a new level “less than \$19,999/year” was created. During phase 4, all source datasets were merged and uploaded into an Oracle database residing on a dedicated secure server.

Assessments—Most variables, especially sociodemographic variables (e.g., age, education, and household income), were obtained from study-specific questionnaires. A few clinical variables (e.g., lifetime history of depression and alcohol use) were from study-

adopted, widely used assessments, including the Structured Clinical Interview for DSM-IV (SCID-I) (First et al. 2002) and the TWEAK alcohol screening test (TWEAK) (Russell 1994). The SCID-I is a clinician-administered semi-structured interview for DSM-IV Axis I disorders. It includes an open-ended overview and diagnostic modules. From this, clinical history can be extracted, including information about lifetime and current treatment and diagnoses. The TWEAK is a five question self-report questionnaire that assesses for harmful drinking habits during pregnancy. A total score greater than two indicates that a woman is likely to be engaged in at risk drinking.

Data analysis

Chi-square tests (X^2) for categorical, t -test for continuous, and Wilcoxon rank-sum test for ordinal variables were calculated to assess systematic differences between (1) EDS and NDS groups and (2) P-EDS and D-EDS groups. For these tests, the independent variable was “group,” and a Bonferroni correction was applied to correct for multiple comparisons. Women with undetermined eligibility were not included in between-group comparisons. At risk or study controls also were not included in between-group comparisons because, although they had elevated depressive symptoms, their agreement to research participation was essentially an agreement to treatment-as-usual.

To compare type and severity of depressive symptoms endorsed by the P-EDS and D-EDS groups, the EPDS items were analyzed. The EPDS was selected between the two DSQs because it was the most shared DSQ among the largest number of participants. To compare type of depressive symptoms endorsed, an indicator variable was created to denote the absence (EPDS item score=0) or presence (EPDS item score = 1) of each symptom. Then, a X^2 for each item was calculated. To compare severity of depressive symptoms endorsed, a Mantel–Haenszel X^2 test for trend for each item was calculated. Severity reflected EPDS coding (i.e., 0, 1...3). In all tests, the independent variable was “group.” A Bonferroni correction was not applied to correct for multiple comparisons since the EPDS items are highly correlated, thus violating an assumption of this correction procedure. All analyses were conducted in SAS, version 9.2 (SAS Institute Inc, Cary, NC).

Results

As depicted in Fig. 1, 2.0 % ($n=186$) of pregnant women did not complete the DSQ that was given during their prenatal care appointment. Based on the source study and integrated dataset’s inclusion criteria, 71.0 % (6,472) of women were identified as NDS because their DSQ scores indicated no-to-minimal depressive symptoms and 23.9 % (2,176) of women were identified as EDS because their DSQ scores indicated elevated levels of depressive symptoms. Because of an incomplete research record, 5.1 % (464) of women had undetermined eligibility. Among women who were identified as EDS, 29.6 % (644) women were identified as P-EDS because they enrolled in a source study, 47.6 % (1,036) as D-EDS because they did not participate in a source study, and 22.8 % (496) as neither because they were had elevated levels of depressive symptoms but were participating in a source study as at risk or study controls.

Table 2 details the sociodemographic and clinical differences between the NDS and EDS groups. Compared to pregnant women in the NDS group, pregnant women in the EDS group were significantly younger, less educated, less likely to be living with spouse or partner, and more likely to be using tobacco and alcohol during pregnancy. The EDS group was significantly more likely than was the NDS group to identify their race as black and report not working or working less than part-time, an annual household income less than \$20,000, more children, poorer physical health, current involvement in mental health treatment for depression, and a history of recurrent depression.

Table 3 shows the sociodemographic and clinical differences between the P-EDS and D-EDS groups. There was only one significant difference: compared to women in the D-EDS group, women in the P-EDS group were at an earlier stage in pregnancy.

There were significant differences between P-EDS and DEDS groups in type or severity of depressive symptoms endorsed on the EPDS. The P-EDS group was significantly more likely than was the D-EDS group to endorse EPDS item 1: ability to laugh, $Z=2.32$, $p=0.02$ and EPDS item 2: ability to enjoy, $Z=1.95$, $p=0.05$. They also were more likely to rate EPDS item 1: ability to laugh, higher than the D-EDS group, indicating greater symptom severity, $X^2(3, N=779)=11.73$, $p=0.01$. There were no other significant differences for EPDS items endorsed or severity rated.

Discussion

Many high impact public health questions about perinatal depression remain unanswered due, in part, to difficulties recruiting statistically sufficient numbers and diverse characteristics of women for participation in depression research. Overall, our experiences establishing the WMHIP and pooling data among WMHIP investigators suggest that it is reasonable to address the problem of insufficient statistical power with post hoc data standardization, integration, and collaborative data sharing (National Advisory Mental Health Council's Workgroup NAMHC 2010).

With a combined, large dataset, we demonstrated that research recruitment strategies embedded within women's routine obstetrical care visits appear to not only be acceptable to women but also minimize selection bias and produce an externally valid study cohort. Almost 98 % of women completed their routine DSQ. Among eligible pregnant women (i.e., those experiencing elevated levels of depressive symptoms), those who did and did not participate in research were statistically similar on all but one sociodemographic and clinical variable (week of pregnancy) and the majority of depressive symptom type and severity variables. Thus, it appears that the sample comprising the integrated dataset is representative of the targeted population—obstetric patients with elevated depressive symptoms. Importantly, this finding contributes to the field by providing evidence that findings from perinatal depression studies, which recruit, based on depression status in obstetrics settings, are applicable to the typical obstetrics patient. Therefore, research using these recruitment methods is a reasonable foundation for clinical care recommendations and guidelines (Khan et al. 2005; Okuda et al. 2010). Future studies could add to this finding by investigating in

greater depth possible sociodemographic biases in those who seek care in obstetrics settings and those who participate in research.

Almost 24% of the pregnant women screened experienced elevated depressive symptoms. This number is well within the range of published prevalence estimates (Gaynes et al. 2005). It is important to note, however, that although the cut point (≥ 10 on the EPDS) has excellent sensitivity (i.e., women suffering are identified the true positive rate) and casts a wide net for detection of illness, it has only good specificity (i.e., women who are not suffering are not identified the true negative rate) (Novick and Flynn 2013).

As expected, pregnant women with elevated depressive symptoms were significantly different from women without elevated depressive symptoms. These women differed in their personal history of depression and on differences associated with social disadvantage, including lower educational attainment, low socioeconomic and employment status, and decreased likelihood of living with a partner or spouse. These between-group distinctions are more likely to reflect the inherent differences associated with risk for and development of prenatal and postnatal depression than to represent differences created by threats to validity, including sample selection, participant enrollment, or attrition (National Research Council and Institute of Medicine NRC 2009; US Department of Health and Human Services Agency US DHHS for Healthcare Research and Quality UDHS AHRQ 2011).

Compared to pregnant women with elevated depressive symptoms who did not participate in a research project, participants were at an earlier stage in pregnancy. They were more likely to endorse impaired “ability to laugh” and “enjoy oneself” and endorse at greater severity impaired “ability to laugh.” One possibility is that women declined participation because their elevated DSQ scores did not reflect depressive symptoms but rather increased physical symptoms and changes associated with the end of the second trimester. Another possibility is that women participated because they were experiencing depressive symptoms earlier in pregnancy and/ or with almost half of their pregnancy remaining. That is, although participants and nonparticipants had statistically similar levels of depressive symptoms at screening, participants had more time to be depressed during pregnancy, and thus, more open to meeting with the research staff and enrolling. Similarly, based on the EPDS differences, participants may have been subjectively more depressed and/or more aware of their depression than nonparticipants, and thus, again more open to meeting with the research staff and enrolling. Essentially, based on the analyses, the typical profile of women who participated in a minimal risk perinatal depression research project were pregnant women who were halfway through their second trimester and reporting, “I can’t laugh, I really can’t laugh, and I can’t enjoy myself.” Future studies should specifically evaluate whether study participation is affected by the timing, duration, or specific symptomatology experienced.

One potential limitation of these findings is the nature of the source studies. All source studies had minimal inclusion and exclusion criteria and favorable risk profiles. Only two source studies involved interventions, and these interventions were non-pharmacological and psychosocial in nature. Therefore, these findings may not be generalizable to investigations involving moderate to high risk or pharmacological agents. Indeed, it is likely that selection

bias is related not only to eligibility criteria but also potential risk associated with research participation, especially for pregnant or breastfeeding women.

Conclusion

Individual studies focusing on perinatal mental health are often limited by relatively small and homogeneous samples. The data integration strategy developed by the WMHIP at the University of Michigan Depression Center allows for expansion of perinatal research questions that require a large sample size to achieve sufficient statistical power. These include investigations of low base rate psychiatric disorders and comprehensive geo-sociodemographic subgroup comparisons.

Our findings suggest that recruitment for minimal risk depression research during prenatal care appointments in obstetric settings appears to be a valid strategy associated with minimal threats to external validity—pregnant women who qualify and enroll in research projects appear to be representative of the larger, targeted population of obstetric patients experiencing depressive symptoms. Additional data integration projects are warranted, especially to replicate our findings and to evaluate the external validity of study cohorts participating in perinatal depression research involving non-obstetric patients, moderate risk, and/or somatic interventions.

Acknowledgments

The Women's Mental Health and Infants Program (WMHIP) was established with support from the Department of Psychiatry at University of Michigan. Original research was supported with grants from the National Institute of Mental Health MH065062 and MH063880, Psychiatry Research Council, and Rachel Upjohn Scholars Award, University of Michigan Depression Center. The authors gratefully acknowledge feedback on a previous draft from Cynthia Battle.

References

- Agency for Healthcare Research and Quality (AHRQ) (2012) The efficacy and safety of screening for postpartum depression. Evidence-based practice center systematic review protocol
- Brody B, Leon AC, Kocsis JH (2011) Antidepressant clinical trials and subject recruitment: just who are symptomatic volunteers? *Am J Psychiatry* 168:1245–1247 [PubMed: 22193668]
- Cox JL (1983) Postnatal depression: A comparison of African and Scottish women. *Soc Psychiatry* 8:25–28
- First MB, Spitzer RL, Gibbon M, Williams JBW (2002) Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition. (SCID-I/P) New York: Biometrics Research, New York State Psychiatric Institute
- Flynn HA, O'mahen HA, Massey L, Marcus S (2006) The impact of a brief obstetrics clinic-based intervention on treatment use for perinatal depression. *J Womens Health* 15:1195–1204
- Gaynes BN, Gavin N, Meltzer-Brody S, et al. (2005) Perinatal depression: prevalence, screening accuracy, and screening outcomes. Agency for Healthcare Research and Quality Report, U.S. Department of Health and Human Services
- Khan AY, Preskorn SH, Baker B (2005) Effect of study criteria on recruitment and generalizability of the results. *J Clin Psychopharmacol* 25:271–275 [PubMed: 15876909]
- Marcus SM, Flynn HA, Blow FC, Barry KL (2003) Depressive symptoms among pregnant women screened in obstetrics settings. *J Women Health Gen B* 12:373–380

- Marcus S, Lopez JF, McDonough S, MacKenzie MJ, Flynn H, Neal CRJ, Gahagan S, Volling B, Kaciroti N, Vazquez DM (2011) Depressive symptoms during pregnancy: impact on neuroendocrine and neonatal outcomes. *Infant Behav Dev* 34:26–34 [PubMed: 21035873]
- National Advisory Mental Health Council's Workgroup (NAMHC) (2010) From discovery to cure: accelerating the development of new and personalized interventions for mental illness. National Institute of Mental Health
- National Research Council and Institute of Medicine (NRC) (2009) Depression in parents, parenting, and children: opportunities to improve identification, treatment, and prevention. The National Academies, Washington, DC
- Novick D, Flynn H (2013) Psychiatric symptoms during and following pregnancy: using psycho-education, assessment, and intervention to improve outcomes. In: Spiers MV, Geller PA, Kloss JD (Eds), *Women's Health Psychology*
- Okuda M, Hasin DS, Olfson M et al. (2010) Generalizability of clinical trials for cannabis dependence to community samples. *Drug Alcohol Depend* 111:177–181 [PubMed: 20537813]
- O'Mahen HA, Flynn HA (2008) (2008). Preferences and perceived barriers to treatment for depression during the perinatal period. *J Womens Health* 17:1301–1309
- Preskorn SH (2003) Relating clinical trials to psychiatric practice: the gap between the usual patient in registration trials and in practice. *J Psychiatr Pract* 9:455–461 [PubMed: 15985968]
- Radloff LS (1977) The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1: 385–401
- Robinson DS, Rickels K (2000) Concerns about clinical drug trials. *J Clin Psychopharmacol* 20:593–596 [PubMed: 11106129]
- Russell M (1994) New assessment tools for drinking in pregnancy: T-ACE, TWEAK, and others. *Alcohol Res Health* 18:55–61
- Stirman S, DeRubeis R, Crits-Christoph P, Brody P (2003) Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *J Consult Clin Psychol* 71:963–972 [PubMed: 14622071]
- U.S. Department of Health and Human Services Agency (US DHHS) for Healthcare Research and Quality (UDHHS AHRQ) (2011) Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evidence Report/Technology Assessment
- Yonkers KA, Vigod S, Ross LE (2011) Diagnosis, pathophysiology and management of mood disorders in pregnant and postpartum women. *Obstet Gynecol* 117:61–77 [PubMed: 21173645]
- Zimmerman M, Posternak MA, Chelminski I (2004) Generalizability of antidepressant efficacy trials. *Essent Psychopharmacol* 6:45–58 [PubMed: 15612653]

Fig. 1.

Classification of research participants

* These participants were not included in any statistical analysis.

DSQ = Depression Screening Questionnaire

Table 1

Details of source studies

Source Study	Design	Primary Aim	Criteria ^f	Procedure	Screened	Eligible	Participated
Understanding mental health treatment use in pregnant and postpartum women ^a	A	Identify and understand common barriers to mental health treatment for women with antenatal depression	Pregnant EPDS 10	Questionnaires on one occasion	465	152	104
Connecting women with care ^b	I	Assess the effect of motivational interviewing (MI) on rates of treatment access and engagement	Pregnant EPDS 12 Not receiving psychotherapy for mood-related issues Receiving ongoing prenatal care with UMHS	Interview and questionnaires on two occasions Random assignment to either two sessions of MI or TAU (psychoeducation and referral)	N/A	N/A	61
Internet depression psychoeducation for pregnant women ^b	I	Assess the effect of an internet-based psycho-education (ED) program on rates of treatment access and engagement	Pregnant EPDS 10	Questionnaires on three occasions Random assignment to either the ED program or TAU (psycho-education and referral from their OB clinic staff)	512	180	78
Screening outcomes in obstetrics ^c	A	Assess health providers' detection of depression during women's prenatal care appointments	Pregnant <12 weeks EPDS 10 Receiving ongoing prenatal care with UMHS	Questionnaires on three occasions	1,785	284	96
The health in pregnancy project ^d	A	Investigate the prevalence of antenatal depression symptoms, alcohol use during pregnancy, and associated risk factors	Pregnant CES-D 16 Receiving ongoing prenatal care with UMHS	Questionnaires on three occasions	4,080	840	234
Perinatal attachment and cortisol study ^e	A	Describe the impact of depressive symptoms during pregnancy on the limbic-hypothalamic-pituitary-axis (LHPA) of women and their infants	Pregnant <28 weeks EPDS 10 Not receiving ongoing psychotherapy Does not meet criteria for current major depressive, bipolar, psychotic, or substance disorder	Questionnaires on 14 occasions Mother and infant saliva and blood samples In-home observations Opportunity to receive free psychotherapy	2,207	619	71

A assessment study, I intervention study, TAU treatment as usual, EPDS Edinburgh Postnatal Depression Screen, CES-D Center for Epidemiologic Studies Depression Scale

^aO'Mahen and Flynn 2008

^bManuscript in preparation

^cFlynn et al. 2006

^dMarcus et al. 2003

^eMarcus et al. 2011

^fInclusion criteria for all studies included fluency in English

Table 2
Differences in sociodemographic and clinical characteristics between pregnant women with NDS and EDS

Variable	NDS (n =6,472)		EDS (n =1,680)		Corrected p value	Effect size
	Mean (SD)	n (%)	Mean (SD)	n (%)		
Age	28.93 (5.56)		26.71 (5.79)		<0.0001	0.39
Race					<0.0001	0.17
White	3,512 (73.70 %)		893 (62.01 %)			
Black	678 (14.23 %)		408 (28.33 %)			
Hispanic	122 (2.56 %)		37 (2.57 %)			
Asian	315 (6.61 %)		49 (3.40 %)			
Other	138 (2.90 %)		53 (3.69 %)			
Married/living with partner	4219 (86.14 %)		994 (68.88 %)		<0.0001	0.19
Number of children	1.00 (1.15)		1.36 (1.46)		0.0003	0.28
Education					<0.0001	0.24
High school diploma or less	1,195 (24.65 %)		660 (45.4 %)			
College graduate	2,415 (49.82 %)		623 (42.85 %)			
Beyond college graduate	1,237 (25.52 %)		171 (11.76 %)			
Employed at least part-time	2,950 (60.36 %)		713 (49.04 %)		<0.0001	0.1
Household income					<0.0001	0.24
Less than \$19,999	349 (21.35 %)		255 (44.35 %)			
\$20,000 to \$59,999	501 (30.64 %)		164 (28.52 %)			
\$60,000 or more	785 (48.01 %)		156 (27.13 %)			
Weeks in pregnancy	24.35 (10.18)		23.45 (10.25)		n.s.	
Self-reported physical health					<0.0001	0.25
Excellent	1,303 (26.47 %)		177 (12.69 %)			
Very good	2,320 (47.13 %)		505 (36.20 %)			
Good	1,179 (23.95 %)		561 (40.22 %)			
Fair	116 (2.36 %)		137 (9.82 %)			
Poor	5 (0.10 %)		15 (1.08 %)			
Using tobacco	400 (11.49 %)		386 (31.72 %)		<0.0001	0.24
TWEAK score	0.40 (0.97)		0.70 (1.34)		<0.0001	0.26

Variable	NDS (<i>n</i> =6,472) Mean (SD) or <i>n</i> (%)	EDS (<i>n</i> =1,680) Mean (SD) or <i>n</i> (%)	Corrected <i>p</i> value	Effect size
Depression treatment at screening	176 (4.32 %)	153 (15.77 %)	<0.0001	0.18
History of recurrent depression (episodes ≥ 2)	119 (19.01 %)	311 (73.87 %)	<0.0001	0.57
History of postpartum depression	152 (15.62 %)	37 (26.81 %)	n.s.	
EPDS total score	5.16 (3.99)	14.22 (3.85)	<0.0001	2.31
CES-D total score	6.64 (4.15)	24.29 (8.09)	<0.0001	2.88

Table 3

Differences in sociodemographic and clinical characteristics between pregnant women with elevated depressive symptoms who did (P-EDS) and did not (D-EDS) participate in perinatal depression research

Variable	D-EDS (<i>n</i> =1036) Mean (SD) or <i>n</i> (%)	P-EDS (<i>n</i> =644) Mean (SD) or <i>n</i> (%)	Corrected <i>p</i> value	Effect size
Age	26.67 (5.76)	26.78 (5.84)	n.s.	–
Race			n.s.	–
White	529 (64.67 %)	364 (58.52 %)		
Black	225 (27.51 %)	183 (29.42 %)		
Hispanic	17 (2.08 %)	20 (3.22 %)		
Asian	25 (3.06 %)	24 (3.86 %)		
Other	22 (2.69 %)	31 (4.98 %)		
Married/living with partner	571 (68.55 %)	423 (69.34 %)	n.s.	–
Number of children	1.30 (1.26)	1.41 (1.61)	n.s.	–
Education			n.s.	–
High school diploma or less	399 (47.67 %)	261 (42.3 %)		
College graduate	328 (38.19 %)	295 (47.81 %)		
Beyond college graduate	110 (13.14 %)	61 (9.89 %)		
Employed at least part-time	400 (47.56 %)	313 (51.06 %)	n.s.	–
Household income			n.s.	–
Less than \$19,999	105 (48.17 %)	150 (42.02 %)		
\$20,000 to \$59,999	58 (26.61 %)	106 (29.69 %)		
\$60,000 or more	55 (25.23 %)	101 (28.29 %)		
Weeks in pregnancy	24.68 (10.69)	21.15 (8.94)	<0.0001	0.36
Self-reported physical health			n.s.	–
Excellent	107 (12.69 %)	70 (12.68 %)		
Very good	309 (36.65 %)	196 (35.51 %)		
Good	346 (41.04 %)	215 (38.95 %)		
Fair	73 (8.66 %)	64 (11.59 %)		
Poor	8 (0.95 %)	7 (1.27 %)		
Using tobacco	213 (30.60 %)	173 (33.21 %)	n.s.	–

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Variable	D-EDS (<i>n</i> =1036)		P-EDS (<i>n</i> =644)		Corrected <i>p</i> value	Effect size
	Mean (SD) or <i>n</i> (%)	Mean (SD) or <i>n</i> (%)	Mean (SD) or <i>n</i> (%)	Mean (SD) or <i>n</i> (%)		
TWEAK score	0.73 (1.36)		0.66 (1.32)		n.s.	—
Depression treatment at screening	91 (13.48 %)		62 (21.02 %)		n.s.	—
History of recurrent depression (episodes ≥ 2)	111 (76.56 %)		200 (72.46 %)		n.s.	—
History of postpartum depression	20 (24.10 %)		17 (30.91 %)		n.s.	—
EPDS total score	13.91 (3.84)		14.61 (3.84)		n.s.	—
CES-D total score	23.88 (8.01)		25.34 (8.22)		n.s.	—