COMPARING EFFECTIVENESS OF TREATING DEPRESSION WITH & WITHOUT COMORBIDITY TO IMPROVE FETAL HEALTH

STUDY PROTOCOL

12/1/2013

CLINICAL TRIALS REGISTRATION NUMBER: NCT02371356

STUDY PROTOCOL

.

Title: Comparing Effectiveness of Treating Depression With & Without Comorbidity to Improve Fetal Health

Short Title: Maternal Depression and Its Treatment in Relation to Fetal Health

Sponsor: Patient Centered Outcomes Research Institute (PCORI)

IRB Number: CN-13-1650-H

Protocol version: 1

Study Principal Investigator: De-Kun Li, MD, PhD
Senior Research Scientist
Division of Research
Kaiser Foundation Research Institute
Kaiser Permanente
2000 Broadway
Oakland, CA 94612
510-891-3755
510-891-3761 (fax)
De-Kun.Li@kp.org

TABLE OF CONTENTS

ABSTRACT

Depression during pregnancy is prevalent (15-20%) and results in adverse fetal outcomes including preterm delivery (PTD) and low birthweight (LBW). Currently, significant confusion exists about unknown risk–benefit profiles of various treatments during pregnancy. We propose to conduct a two-stage prospective cohort study, using KPNC's peripartum depression screening program, to determine if treating depression in pregnancy is effective in improving fetal outcomes, and which treatment is most effective: pharmacotherapy, psychotherapy or a combination. Risk-benefit of treatments will be examined separately for two depression types: those with depression only and those with psychiatric co-morbidities to evaluate differences in treatment effectiveness. Four cohorts with different treatment options will be formed within each depression type (with or without co-morbidity): (A) "Antidepressant only": depressed and use only antidepressants; (B) "Psychotherapy only": depressed and receive psychotherapy only; (C) "Combination therapy": depressed and receive antidepressants and psychotherapy; (D)"Untreated depression": depressed and receive no treatment. Eight cohorts will be formed. We will also form cohort (E) "No depression": screen negative and receive no treatment, for baseline comparison. Information on depression treatment and PTD & LBW will be available for all 88,000 women in the stage-one sample. We will form stage-two sub-cohorts by randomly selecting 400 from each cohort (total 3,600) and interview them to obtain detailed information on treatment compliance and confounders. Comparison of Cohort A, B & C to D, respectively, will determine if treating depression is effective. Pair-wise comparisons among Cohorts A, B & C will determine comparative effectiveness of treatments. Comparison of Cohort D to E provides baseline fetal risks of untreated depression. Findings will provide answers to pressing questions of how to treat depression in pregnancy and which treatment is most effective with the best risk-benefit profile in improving fetal outcomes, thereby reducing PTD or LBW, infant mortality and morbidity, and medical costs.

1 BACKGROUND INFORMATION AND RATIONALE

A.1. Impact of the Condition on the Health of Individuals and Populations

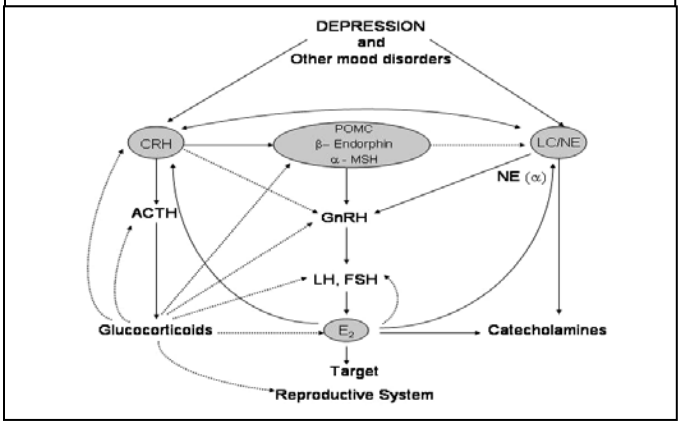
A.a.1. Impact of maternal depression during pregnancy

More than 20% of pregnant women may have a depressive disorder or symptoms during pregnancy [1-3]. However, it is well-recognized that depression during pregnancy is significantly underappreciated and under-recognized, thus under-diagnosed [1, 4-9]. While there have been no good data on the degree of under-diagnosis of depression during pregnancy, one of the studies conducted in the Kaiser Permanente Northern California (KPNC) population suggested a detection rate of only 37% for postpartum depression [10]. This detection rate is comparable to that reported in the literature[11-13]. Various studies including a study conducted by the Principal Investigator among the KPNC population have shown that prenatal depression is likely to affect about 20% of pregnant women[3 9, 14-16]. Therefore, depression during pregnancy is prevalent and has significant clinical and public health importance. Emerging evidence has shown that depression in pregnancy increases the risk of preterm delivery (PTD) and low birthweight (LBW).

A.a.2. Impact of PTD and LBW

PTD, defined as birth before 37 completed weeks of gestation, is the leading cause of perinatal mortality and morbidity. It accounts for about 30% of early neonatal deaths worldwide [17, 18]. It is also the leading cause of congenital neurological disabilities including cerebral palsy, blindness and deafness and learning disability in the U.S. and most developed countries [19-22]. In addition, it is the major cause of admission to the neonatal intensive care unit and a significant contributor to medical expenditure for infants. The costs associated with PTD amount to more than $26 billion each year [23-25]. Such impact on infant health and on staggering medical and economic costs makes PTD one of the most serious challenges for clinicians and biomedical researchers today and has become a global crisis [26, 27]. Thus, reducing PTD and LBW could lead to a significant reduction in medical care costs and to improved newborn health. However, despite decades of research, the incidence of PTD has not been reduced; instead, it has been steadily increasing [20, 28-30]. In the US, about 12% of all births (>half a million births) are the result of PTD each year.

Figure 1. Potential mechanistic pathways between depression and adverse pregnancy outcomes



A. a.3. Prenatal depression and the risk of PTD and LBW

While the causes of PTD remain largely unknown, which explains the lack of progress in reducing the incidence of PTD, emerging evidence, including a study conducted by the PI, has linked depression during pregnancy to PTD risk[3]. In addition, a growing body of literature has provided underlying mechanisms for the observed link between depression during pregnancy and PTD risk. Through the direct effect of hypothalamic-pituitary-adrenal (HPA) axis and its indirect effect by the interaction between HPA and HPG (hypothalamic-pituitary-gonadal) axes, psychopathological factors could

have a profound impact on the female reproductive system and pregnancy outcomes including PTD (Figure 1) [31-35].

A. a.4. Epidemiological evidence of fetal risk with untreated depression during pregnancy

Emerging epidemiological studies including one conducted by the PI (see Preliminary Studies) have shown that the presence of significant and severe depressive symptoms during pregnancy increased the risk of PTD [36-39]. Given that only a fraction of depressed pregnant women were clinically diagnosed [10, 40], all studies used population-based screening instruments for depression symptoms such as the Patient Health Questionnaire (PHQ-9) or Center for Epidemiological Studies Depression Scale (CES-D). These instruments have been shown to be highly correlated with clinical diagnoses with 85-90% sensitivity and specificity [39, 41, 42]. Severe depressive symptoms have also been demonstrated to be a good indicator for clinical diagnosis of depression [42-44]. Women with significant depressive symptoms, usually ascertained in the second and third trimesters, had increased risk of PTD with a relative risk ranging from 1.9 to 6.8 [36-38]. A similar association was also reported for LBW [45-47].

A. a.5. Treatment of depression during pregnancy and its effect on PTD and LBW

Given that depression during pregnancy is an important risk factor for both PTD and LBW, it would seem logical to treat depression during pregnancy to reduce these risks. However, two obstacles have made this seemingly straightforward decision complicated, especially for pharmacological treatment of depression (i.e. antidepressants). First, it is well-recognized that prenatal depression is significantly underappreciated and under-diagnosed [1, 4-9]. Without first identifying pregnant women with depression, evaluating treatment effectiveness would likely not be valid. Second, in addition to treatment effect on depression, the safety concerns of antidepressant use during pregnancy has been raised. Studies have shown that use of antidepressants during pregnancy itself may increase the risk of PTD and other adverse outcomes[48-51]. However, most of those studies compared users of antidepressants to non-users who were mostly pregnant women without underlying depression. Due to a lack of universal screening for depression during pregnancy, those studies were not able to identify a relevant comparison group of pregnant women with untreated depression. Thus, those findings were unable to disentangle the pharmacologic effect of antidepressant use during pregnancy from the effect of underlying depression, confounding by indication. Currently, the risk-benefit balance of treating depression during pregnancy remains uncertain, and comparative effectiveness studies to examine the risk-benefit profile of depression treatment during pregnancy in relation to fetal risk are urgently needed to provide evidence-based treatment decisions.

A. a.5.1. Psychotherapy treatment and the risk of PTD

While psychotherapy is not expected to increase the risk of PTD, evidence for a beneficial effect of psychotherapy treatment of depression during pregnancy on reducing the risk of PTD has been absent largely due to a lack of examination of such an effect. A recent review was able to identify only one small study (38 subjects) that examined the effect of psychotherapy on treatment of depression, but not its subsequent impact on the risk of PTD [52, 53]. Essentially, the effect of psychotherapy treatment of depression during pregnancy on the risk of PTD has not been examined. The effect of psychotherapy for treatment of depression during pregnancy including timing and duration of such psychotherapies on the risk of PTD is currently unknown.

A. a.5.2. Antidepressants and the risk of PTD and LBW

A difficult challenge in deciding treatment options for depression during pregnancy is the potential adverse effects of antidepressant use on fetal risk including PTD and LBW. While their potential benefit to mitigate depression during pregnancy is well-known [54], the potential adverse pharmacological effects of antidepressant use during pregnancy on fetal risk have been raised by a limited number of studies. While research findings were not consistent, a few studies reported an increased risk of PTD and, to a lesser extent, LBW among women who used antidepressants, both SSRI (selective serotonin reuptake inhibitors) and non-SSRI antidepressants [55-62]. However, many of these studies had methodological problems that are common in studying pharmacological effects [1]. Because the comparison groups were usually women with neither depression nor antidepressant use, it is difficult to disentangle whether the high risk of PTD among women who used antidepressants was due to the pharmacological effect of antidepressants or the effect of the underlying depression itself. A more valid comparison would be a cohort of women with depression, but who did not use antidepressants (i.e., untreated controls).

Some earlier studies did not observe an increased risk of PTD with antidepressant use (mainly SSRIs) [63] and one study even reported a decreased risk of PTD among women who used antidepressants [64]. One study reported an increased risk of PTD associated with SSRIs although the number of women who had both depression and used SSRIs was small (n=55)[62]. Therefore, the effect of antidepressant use during pregnancy on the risk of PTD remains largely uncertain. The crucial issue in studying the effect of antidepressants remains to identify a comparison group consisting of women with depression who do not use antidepressants (untreated controls). Our proposed study will provide exactly such an appropriate comparison group thanks to the KPNC universal peripartum depression screening program. Furthermore, the proposed study will be able to examine comparative effectiveness of various treatment regimens among pregnant women with similar underlying depression *severity* thanks to universal depression screening.

A.a.6. Depression with and without comorbidity

Depression with and without concurrent other psychiatric disorders may be different disease entities with different underlying pathophysiological pathways [65, 66]. Depression with psychiatric comorbidity (psychotic depression) has been associated with a greater HPA axis dysregulation than depression without comorbidity [66]. Treatment for these two types of depression has been different [67-69]. Therefore, their impact and treatment effectiveness on fetal risk are likely different. It would be clinically informative and valuable to examine these two types of depression and their treatment effectiveness separately.

In summary, while depression during pregnancy is an important risk factor for adverse pregnancy outcomes, treating depression during pregnancy has proven to be a difficult decision for both clinicians and pregnant women to make due to the uncertain fetal risk of antidepressant use[1, 52, 70]. Both the American College of Obstetricians and Gynecologists (ACOG) and the American Psychiatric Association (APA) have called for more research to demonstrate the comparative effectiveness of treatment choices in reducing fetal risk and to better understand risk-benefit of various treatment therapies[71-73]. Given the potential ethical problems in randomizing pregnant women into treatment regimens with potential fetal risk, the primary concern of pregnant women, carefully designed prospective studies to compare treatment effectiveness for various treatment choices for depression during pregnancy are likely the next best alternative approach. The first step for conducting such a comparative effectiveness study is to have a universal depression

screening to identify women with depressive symptoms, especially those who do not seek diagnosis or treatment (those with untreated depression). ACOG issued a call for "Screening for Depression During and After Pregnancy"[74, 75].

KPNC has implemented a region-wide universal screening program for peripartum depression in over 40 clinical facilities including 16 delivery hospitals, covering racially/ethnically diverse populations. This is likely the nation's first large scale screening program for peripartum depression which screens more than 35,000 pregnant women annually. Thus, KPNC provides a unique setting to conduct the proposed comparative effectiveness study of depression treatment during pregnancy to reduce fetal risk and improve pregnancy outcomes, including head-to-head comparison between pharmacotherapy and psychotherapy.

B.3 Benefit to patients and their clinicians

As stated above, our proposed study will provide answers to important questions that are significant challenges faced by both pregnant women and their clinicians. When making treatment decision during pregnancy, pregnant women and their physicians face the complexity of needing to consider "Treating for Two". The results of our study will provide important evidence to allow pregnant women and their physicians to make informed decisions based on the findings of (1) the risk of PTD and LBW if not treating depression during pregnancy and (2) the benefit of treating depression in reducing the risk of PTD and LBW. Answers to those questions will be valuable to pregnant women and clinicians, and directly address their concerns when making treatment decisions.

2 STUDY OBJECTIVES

Specific Aims: Depression during pregnancy is prevalent with nearly one million pregnant women suffering from it in US annually. Depression, if UNTREATED, has been linked to adverse fetal outcomes including preterm delivery (PTD) and low birthweight (LBW). Currently, significant confusion exists about risks and benefits of treating depression during pregnancy. Due to ethical concerns of randomizing pregnant women into treatment regimens with potential fetal risks, observational studies are likely the best alternative to RCTs, and have the advantage of providing real-life evidence of risks/benefits of depression treatment. Based on an universal depression screening program, we propose to conduct a comparative effectiveness study with a *two-stage* prospective cohort design to determine if treating depression in pregnancy is effective in improving fetal outcomes, and which treatment has the best risk-benefit profile: pharmacotherapy, psychotherapy or a combination. The questions will be examined separately for two depression types, *depression only* and *depression with other psychiatric co-morbidities,* to elucidate possible differences in treatment effectiveness between the two groups. Four cohorts will be formed within each depression type: (A) "Antidepressant only": screen positive for depression and use only antidepressants during pregnancy; (B) "Psychotherapy only": screen positive and receive psychotherapy only; (C) "Combination therapy": screen positive and receive both antidepressants and psychotherapy; (D)"Untreated depression": screen positive and receive no treatment. We will also form a cohort (E) "No depression": screen negative and receive no treatment, for baseline comparison. Information on depression, treatment and PTD & LBW will be available for all 88,000 women (stage-one sample). To control for confounding, we will randomly select 400 women from each cohort (total 3,600) for a telephone interview to obtain detailed information on confounding factors. Information from both stages will allow efficient and valid estimates while controlling for confounders. The aims are:

Aim 1. Do women who receive treatment during pregnancy have a lower fetal risk than women with untreated depression? (Cohort A, B & C vs. Cohort D).

Aim 2. Which treatment is more effective in reducing fetal risk? (Pair-wise comparison among Cohort A, B & C, respectively)

Aim 3. Do women with untreated depression have a higher fetal risk than women without depression? (Cohort D vs. E).

Aim 4. Does the treatment effectiveness differ between women with depression only and women with depression and other comorbidity?

Exploratory analyses will be conducted to examine whether treatment effectiveness (Aims 1 & 2) is impacted by baseline depression severity. The aims will be refined with continued input from stakeholders. This will be the first large scale comparative effectiveness study to answer pressing clinical questions with an innovative and efficient two-stage design.

3 STUDY PROCEDURES

3.1 Study Design

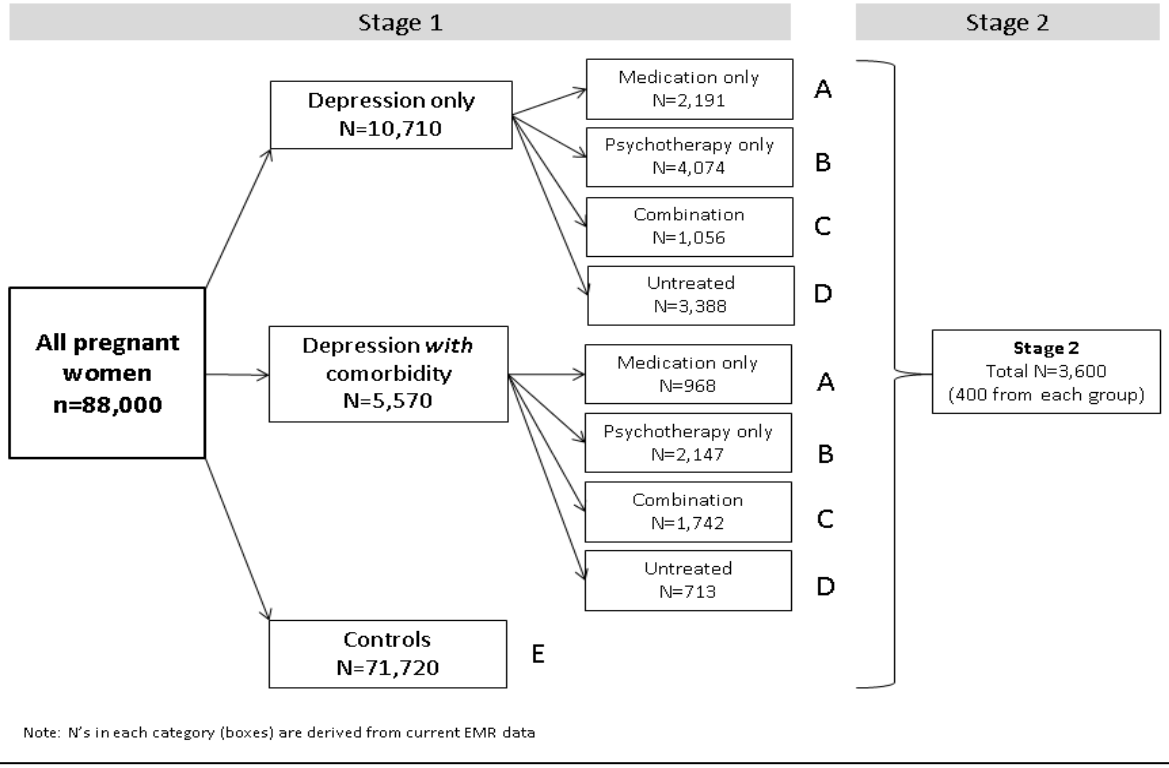
3.1.1 Overview of the Study Design and Approaches

The proposed study will be a population-based prospective cohort study with an efficient two-stage design. This study design is uniquely suited for the KPNC available infrastructure, an advanced and comprehensive EMR system and universal peripartum depression screening program. Given that recalling depression during pregnancy in the presence of adverse pregnancy outcomes (PTD or LBW) is likely biased (recall bias), retrospective study designs like case-control studies are not appropriate for the objectives of the proposed study. Thus, a prospective cohort study design obtaining information on depression and its treatment during pregnancy before pregnancy outcomes is the only appropriate study design for addressing the Specific Aims of the proposed study. In addition, KPNC's unique infrastructure allows for an efficient two-stage prospective cohort study. The KPNC comprehensive EMR which captures the PHQ-9 scores for participants of the universal depression screening allows us to identify information on depression and its treatment status for 88,000 pregnant women expected to be screened. KPNC EMR will also allow us to ascertain pregnancy outcomes, PTD and LBW; thus, we will have information on exposures (depression and its treatment) and outcomes (PTD and LBW) for all 88,000 pregnant women during the study period to form 4 comparison cohorts within each depression type (with or without comorbidity) and one control cohort (a total of nine cohorts). While it is not practical to interview 88,000 women, contacting and interviewing 3,600 (400 from each cohort) in the stage-two sample will allow us to obtain information on confounders. Information on underlying depression severity at baseline allows us to control for confounding by indication and to examine comparative effectiveness of treatment options among women with similar severity of underlying depression. Other psychiatric disorders captured by KPNC EMR will allow us to examine treatment effectiveness separately for two different types of depression. Combining information from both stage-one and stage-two populations allow us to examine the Specific Aims using an appropriate study design (prospective cohort study) with an efficient approach (two-stage design). KPNC's infrastructure makes this unique combination of study designs possible.

3.1.2 Choice of outcomes

PTD, defined as birth before 37 completed weeks of gestation, is the leading cause of perinatal mortality and morbidity as well as a leading cause of congenital neurological disabilities including cerebral palsy, blindness and deafness in the U.S. and most developed countries [19-22]. In addition, it is the major cause of admission to the neonatal intensive care unit and a significant contributor to medical expenditure for infants, amounting to more than $26 billion each year[23-25]. Such impact on infant health and on staggering medical and economic costs makes prematurity one of the most serious challenges for clinicians and researchers today and has become a global crisis according to WHO reports and March of Dimes White Paper [26, 27]. Patients with premature infants have severe physical, emotional, and financial burdens beginning at delivery. Premature birth turns a happy event (having a new child) into a nightmare. Risk factors for prematurity are largely unknown, thus there has not been much progress in reducing premature births largely due to a lack of intervention strategy, which leaves pregnant women to face significant uncertainty throughout pregnancy. Given depression in pregnancy has emerged as an important and potentially modifiable risk factor for prematurity, identifying an effective treatment option for pregnant women with varying type and severity of underlying depression could prove a valuable intervention strategy to reduce prematurity, which could be a significant contribution and help for pregnant women and society in general.

Figure 2. Study Design



Note: N's in each category (boxes) are derived from current EMR data

3.1.3 Data collection

Information on exposures (depression status, and treatment), outcome (PTD & LBW), and other psychiatric disorders (comorbidity) will be ascertained through KPNC EMR data sources for all

88,000 women who have been screened for peripartum depression. Confounders not available from the KPNC EMR will be ascertained during interview of the stage-two subsample of 3,600 women (Figure 2). OBGYNS will inform all pregnant women who screen positive by PHQ-9 about our study during their prenatal visits. Table 1 describes briefly the sources and categories of factors to be ascertained.

Table 1: Sources of Exposures, Outcomes, Confounders and Modifiers (category only)

		Universal Screening	EMR data	Interview
Depression and treatment	Depression and its severity	X	X	
	Psychotherapies	X	X	
	Antidepressant use	X	X	
	Combination therapy	X	X	
	Other psychiatric disorders		X	
Outcome	Gestational Age/PTD, LBW		X	
Potential confounders, mediators and effect modifiers	Risk factors for PTD, LBW		X	X
	History of mental illness		X	X
	Reproductive history		X	X
	Preexisting diabetes or GDM		X	X
	BMI		X	X
	History of depression	X	X	X
	Substance abuse		X	X
	Reproductive and pregnancy history		X	X
	Family and social support		X	X
	Demographic characteristics		X	X
	Family history of mental illness			X

Depression status and severity

Depression status and its severity during pregnancy will be ascertained through PHQ-9 scores from the universal depression screening and clinical diagnosis. All pregnant women who come to the first prenatal visit will be screened for depression using the PHQ-9, which has been validated in many studies as an instrument for screening for depression with high sensitivity (> 88%) and specificity (> 88%) in obstetric patients [103-107]. PHQ-9 was chosen as the screening instrument after consultation with many nationally recognized experts in the field of diagnosis and treatment of peripartum depression balancing out many factors including scientific validity and feasibility for a large scale population-based screening program. As a currently accepted cutoff, PHQ-9 scores (≥10) in combination with evaluation by medical practitioners will be used to determine whether a pregnant woman screens positive for depression. A score above 10 will be used as a measure of the severity of depression. Women who screen positive are referred by clinicians, under the screening protocol, to departments of psychiatry, psychology, or behavioral modification. Information on PHQ-9 scores, referrals, patient's uptake (or decline) of prescribed treatments are also captured by our EMRs. Any clinical diagnosis and treatment of depression for those not captured by the screening will also be identified through EMRs. During the glucola visit (22-24 weeks of gestation), all women will be screened for depression a second time using the same screening process as the first prenatal visit.

Women being treated for depression before the screening will be identified (<2%) and analyzed as a separate group.

Depression severity will also be used for stratification (to control for confounding by indication) when conducting head-to-head comparison of treatment effect among various treatment options.

Depression type (with or without other psychiatric comorbidity) will be determined by identifying concurrent psychiatric disorders from KPNC EMRs including organic psychotic conditions, other psychoses, neurotic and personality disorder and other mental disorders. The expected number of women who have depression only and depression with comorbidity during the study period is listed in Figure 2.

Antidepressant use during pregnancy. Antidepressant use, including types, timing, and duration of use, will be ascertained both through the depression screening and EMR (i.e., pharmacy database). The KPNC pharmacy database captures all inpatient and outpatient prescriptions and dispensations. Because of a large financial incentive to obtain medications at KPNC (in most cases, medications are covered with minimal co-payments), the KPNC pharmacy database has been shown to capture 93-98% of prescription medications of varying types. Compliance of taking antidepressants will also be ascertained during interview.

Psychotherapy. KPNC EMR records any psychotherapy patients receive for depression. The information includes the type of therapies (individual or group), sources of the treatment (psychologist, behavioral modification specialist, or psychologist/social worker) and number of sessions attended.

Outcomes (PTD & LBW): Gestational age at delivery is determined by obstetricians based on multiple sources including ultrasound dating, physical examination, and LMP. Birthweight is measured by medical practitioners at delivery. This information is recorded in the KPNC EMR. In several studies conducted by the PI, more than 99% of participants had reliable information on gestational age and birthweight at delivery in the KPNC EMR. For the remaining <1% that may have delivered outside KPNC hospitals, we are also able to obtain information through searching various databases including outside payment service databases. In addition, KPNC has a well-established Neonatal Minimum Data Set which verifies all NICU admissions for diagnosis and treatments. This database allows ascertainment of information on gestational age and birthweight for deliveries with missing information. We will also ascertain other related information such as indications for PTD and other pregnancy or labor complications for further classification of PTD. A sample of PTD and LBW records will be selected and verified by manual review of medical records for quality control.

Potential confounders. In the stage-two subsample, we will conduct a structured telephone interview to ascertain detailed information on potential confounders that may not be available from the KPNC EMR. The focus will be known and suspected risk factors for PTD and LBW, personal and family history of mental illness, demographic, reproductive and psychosocial factors (e.g., social and family support that may be related to depression, PTD, and LBW), BMI, diseases during the index pregnancy (e.g., GDM), diet, lifetime and prenatal smoking. The PI and his research team have extensive experience ascertaining information on these variables for many of his previous NIH-funded studies. Questionnaires with these questions have been used in many previous

studies. Based on our experience conducting telephone interviews with KPNC pregnant women, a response rate of 80-85% is expected.

4 STUDY DURATION, ENROLLMENT AND SITES

The proposed study will be completed within three years and will involve eligible patients from all Kaiser Permanente Northern California facilities. Start-up will occur during the first three months and will involve obtaining IRB approval, hiring staff, training, conducting focus group interviews, testing data collection materials, developing and revising interview instruments (including the questionnaire) for stage-two data collection of information on confounders, establishing mechanisms for identification of eligible subjects through the EMR. Data collection (including data linkage, establishment of depression screening registry, recruitment and conducting interviews) will be ongoing from month 4 through month 27. Ascertainment of diagnosis and treatment of depression and pregnancy outcomes will continue from month 10 through month 31. Data cleaning and analysis as well as report writing will be carried out during the last 6 months. The timeline for activities involved in the study is as follows:

Months 1-3	Obtain IRB approval. Hire staff. Train interviewers. Set-up tracking database. Develop algorithm for identifying eligible participants. Create interview instruments including questionnaire and instructions, operations manual, and pilot test questionnaire. Conduct focus group interviews.
Months 4-27	Identify and include eligible pregnant women for the stage-one sample. Conduct data linkage and build the *peripartum depression screening registry.* Classify 88,000 women from stage-one sample into treatment categories based on their depression status and treatment regimens. Select a random sample of 400 from each category to form the stage-two subsample. Conduct telephone interviews with stage-two subsamples.
Months 10-31	Ascertain outcomes of interest.
Months 30-35	Clean data, conduct and complete data analyses, and prepare final report.
Months 35-36	Create materials for dissemination of the findings. Dissemination of the findings, which will continue beyond month 36.

4.1 Study Population

4.1.1 Inclusion Criteria

1) Females ≥ age 18

2) Kaiser members

3) Pregnant beyond 20 weeks of gestation during the recruitment period

4) Meeting the definition of one of the following nine cohorts:

Cohort A (depression only): pregnant women who are depressed and use only antidepressants

Cohort A (depression with comorbidity): pregnant women who are depressed and use only antidepressants

Cohort B (depression only): pregnant women who are depressed and receive psychotherapy only

Cohort B (depression with comorbidity): pregnant women who are depressed and receive psychotherapy only

Cohort C (depression only): pregnant women who are depressed and receive antidepressants and psychotherapy

Cohort C (depression with comorbidity): pregnant women who are depressed and receive antidepressants and psychotherapy

Cohort D (depression only): pregnant women who are depressed and receive no treatment

Cohort D (depression with comorbidity): pregnant women who are depressed and receive no treatment

Cohort E: pregnant women who are not depressed (screen negative) and receive no treatment

4.1.2 Exclusion Criteria

1) <18 years old

2) Not Kaiser Permanente members

3) Not pregnant.

4) Not meeting the definition for the nine cohorts (pregnancy and depression/treatment status).

Subjects that do not meet all of the enrollment criteria may not be enrolled.

4.1.3 Populations

Stage I population

All KPNC women aged 18 years (legally allowed to participate in these types of studies at KPNC) or older who are pregnant and who participate in the universal peripartum depression screening will be eligible for the study. We expect that about 88,000 pregnant women will be

eligible during the study recruitment period. If a woman becomes pregnant a second time during the study period (an unlikely event given the short recruitment period of less than three years) and she was already in the study for her previous pregnancy, her second pregnancy will not be included in the study so as to avoid non-independent observations.

All participants in the universal depression screening will be identified and information on screening results, referrals, treatments, pregnancy related conditions and outcomes will be collected through linkage to various data sources of our comprehensive EMR system.

Stage II samples

A sample of 400 pregnant women from each of the nine cohorts previously described will be randomly selected to form Stage II samples. The random selection will be spread evenly throughout the study period so that all of the estimated 88,000 participants in Stage I will have an equal chance of being selected. They will be selected after their inclusion in Stage I with known status of depression and treatment, mostly before the 3rd trimester to avoid interviewing after delivery to reduce participation and recall biases. Any change in exposure status (depression and treatment) after initial assignment will be rectified and finalized during analyses, as we have done in other similar studies with two-stage design (with different exposures in pregnancy).

Participants in Stage II samples will be interviewed to obtain more detailed information on depression and its treatments. The Stage II samples will also allow for direct ascertainment of compliance with prescribed medications. Studies based on pharmacy data alone usually have to make an assumption of 100% compliance with dispensed medications. In this study, we can directly examine the compliance and assess its potential impact on treatment effectiveness. This will be an important strength for this study.

Through the interview, we will also directly ascertain compliance to prescribed medications for treating depression. We will make appropriate adjustment for compliance ratio in data analysis to more accurately estimate the treatment effectiveness after taking into account of non-compliance. In addition, we will also ascertain additional confounders including multivitamin intake during pregnancy and physical activity.

Informed consent will be obtained before the interview. All contact procedures will follow the standard contact and recruitment procedures as required by KPNC's IRB. Based on our experience with similar studies among KPNC pregnant women, a participation rate of 70-75% among eligible pregnant women can be expected. The recruitment will continue until 400 subjects from each of the nine cohort groups have completed the interview.

5 COMPARATORS

Comparators will vary depending on the research questions to be examined in analyses. The overall objective is to use appropriate comparators so that correct results of comparative effectiveness can be obtained after disentangling treatment effect from the effect of underlying depression (confounding by indication), and treatment benefits after balancing treatment benefit against the simultaneous risk of side effect (e.g., use of antidepressants). The following four cohorts (A-D) will be established within each depression type (depression only or depression with comorbidity) and various comparators will be used to answer the research questions in the Specific

Aims. We will also establish a control cohort (E) of women without depression for comparison of baseline risk.

Cohort A: women who screened positive and used antidepressants only. All women who screened positive for depression and were subsequently prescribed antidepressants during pregnancy will form Cohort A. Comparison of this cohort to Cohort D, a comparator for Specific Aim 1, will demonstrate the treatment effect of antidepressant use in reducing the fetal risk after controlling for severity of depressive symptoms and other confounders. Comparisons to Cohort B or C, comparators for Specific Aim 2, will determine which treatment option (antidepressant use, psychotherapy or combination therapy) is more effective in reducing fetal risk or has a better risk-benefit profile (pair-wise comparisons). Results from such comparisons will provide valuable information for clinicians as well as pregnant women to make an informed decision on the treatment options for depression during pregnancy based on research evidence.

Cohort B: women who screened positive and received psychotherapy only. This cohort consists of women who screened positive and subsequently received psychotherapy during pregnancy, but did not take any antidepressants. Comparison to Cohort D will determine if mitigation of depression by psychotherapy reduces fetal risk, after controlling for depression severity. Comparisons of this cohort to antidepressant use (Cohort A) or combination therapy (Cohort C), two separate comparators for Specific Aim 2, will provide head-to-head assessment of comparative effectiveness of treatment regimens.

Cohort C: women who screened positive and used both antidepressants and psychotherapy. This will be a cohort similar to Cohort A and B except that women in this cohort received combination therapies after screening positive. We classify them in a separate cohort to make Cohorts A and B encompass more clearly defined exposures of antidepressants and psychotherapy (pure exposure groups). In addition, this cohort may allow us to examine the potential benefit of combination therapy. Like Cohorts A and B described above, they will be compared to untreated comparator (Cohort D) for Specific Aim 1 and other treatment comparators (Cohort A and B) for Specific Aim 2.

Cohort D: women who screened positive for depression, and did not obtain treatment. Pregnant women whose PHQ-9 score is ≥10 at either of two screening visits (the first prenatal visit and glucola visit) or who were diagnosed as having depression by their medical providers, but did not receive any treatment for depression during pregnancy, will be classified into this cohort. Comparison of this cohort to Cohort E (control cohort), a comparator for Specific Aim 3, will allow us to assess whether untreated depression during pregnancy increases fetal risk (i.e., PTD or LBW) without the interference of treatment effect. We can also examine potential dose-response relationships based on the severity of depression (PHQ-9 scores) and duration of depression (persistence of the symptoms in both screening visits vs. at only one of the visits). Importantly, this cohort can be used as a comparator for Cohorts A, B, and C to examine the effect of depression treatment after controlling for underlying depression and severity (untreated comparators).

Cohort E (controls): women who screened negative for depression in both visits and did not receive treatments for depression. This cohort consists of the largest number of pregnant women, those who have neither depression and concurrent psychiatric comorbidity, nor treatment during pregnancy. This cohort will serve as the comparator for determining the effect of untreated depression on fetal risk (i.e., PTD, LBW). It will also serve as a reference check for other

comparisons. For example, does the benefit of a treatment totally eliminate the risk of PTD associated with depression during pregnancy?

The categorization into each cohort will be finalized at the end of pregnancy based on whether a woman met the criteria throughout the pregnancy. Comparisons of Cohort A, B, C, vs. D respectively, will answer the question of whether treating depression during pregnancy improves fetal outcomes, reducing PTD or LBW associated with untreated depression after controlling for underlying depression severity. Pairwise comparisons among Cohort A, B and C will answer the question of which treatment is most effective in improving fetal outcomes and has the best risk-benefit profile.

6 ANALYTIC METHODS

6.1 General Approach to Analyses

While all eligible pregnant women will be included in the registry after their pregnancies are identified, the follow-up time for all pregnancy outcomes of interest will start at 20 weeks of gestation because, by definition, PTD and LBW among live births occur after 20 weeks of gestation. All pregnancies ending before 37 completed weeks (259 days) will be considered as failure (PTD). As in our previous studies, we will use the Cox proportional hazards regression model, accommodating delayed entry into the cohort[76, 77]. Birthweight for all newborns will be obtained from EMR and analyses for LBW will utilize logistic regression.

The study design can be characterized as two-stage. Unlike typical applications of two-stage sampling in epidemiologic studies, where the stage-two sample is selected on the basis of both exposure (or exposure surrogate) and outcome (e.g. two-stage case-control sampling), only exposure status will be known at the time of stage-two sampling to keep the nature of a cohort study design. However, study outcomes of interest will eventually be measured on all 88,000 stage-one units, and there is much to be gained by utilizing all available information in estimation of associations of interest (see Sample Size and Statistical Power section).

Regression parameter estimates and associated standard errors will be obtained via semiparametric maximum likelihood [78], utilizing exposure and outcome measurements on all 88,000 stage-one units in addition to confounder/effect modifier measurements on the subset sample at stage-two (N=3,600). Scott and Wild presented a unified method for fitting arbitrary regression models to a large class of missing data and/or response selective sampling problems. These applications are generally characterized as: 1) a set of easily-obtainable variables is measured on a sample of N individuals; one or more of these variables are to be used as explanatory in the regression model (e.g. depression and treatment status), or are informative surrogates for "expensive" variables to be measured on a subsample; 2) the response variable is obtained for all N individuals (e.g. PTD, LBW); and, 3) a set of "expensive" explanatory variables (e.g. potential confounders/effect modifiers) are measured on a subsample. Software (R language) has been developed and available from the authors, with functions for various regression models including binary regression, linear regression for continuous response, and clustered binary data. The associated software documentation provides details on key parts of the system enabling users to implement new regression models (e.g. Cox). We will utilize logistic and Cox regression to address the study aims, as described below.

Each set of analyses addressing Aims 1, 2 and 3, outlined below, will be conducted in samples of women with and without other psychiatric comorbidities (Aim 4).

Aim 1. Treatment effect compared to untreated depression (Cohort A, B & C vs. Cohort D)

Point and interval estimates of the relative hazard of PTD associated with each treatment modality (A, B, C), relative to no treatment (D) will be calculated via semiparametric maximum likelihood estimates of Cox proportional hazards regression parameters, with adjustment for potential confounders, as described above. In determining potential confounders in our regression models, we will first examine risk factors that may be plausibly linked to depression and PTD. Inclusion of potential confounding factors in our final regression models will be evaluated based on comparison of adjusted and unadjusted relative hazard ratios [79]. Departures from model assumptions will be assessed via diagnostic plots of weighted residuals and tests for interaction between exposure and time (gestational age). The assessment of confounding variables for inclusion in regression models and testing of model assumptions will be performed in the stage-two sample using standard software for fitting Cox regression models. After determining the appropriate set of confounders and that there are no departures from model assumptions, the stage-one information on primary exposure and outcome will be incorporated using the semiparametric maximum likelihood estimator. This process applies to analyses of the remaining aims.

These comparisons will allow us to assess possible treatment effects on reducing PTD through both pharmacotherapy and psychotherapy. When compared to untreated (Cohort D), the comparison will be made after controlling for underlying severity of depression as measured by PHQ-9 scores or clinical diagnosis of severity before treatment. Table 2 shows that depression severity is relatively comparable among comparison groups, allowing control for confounding by indication (underlying depression severity, a unique strength of the proposed study due to the availability of PHQ-9 scores through the universal depression screening).

We will also examine the possible dose-response in the association between antidepressant use and PTD by examining dosage of use, duration of the use, and timing of effect (single vs. multiple trimesters), in the context of time-dependent covariates. Similarly, we will also assess any dose-response relationship in the association between psychotherapy and PTD (the number of sessions received) and the effect of timing of treatment, also both as time-dependent covariates. If sample sizes allow, we will also examine the effect of combination of dose and timing (e.g., dose-response during early and late pregnancy). In addition, we will examine difference in the type of antidepressants (SSRI vs. non-SSRI).

Analyses of LBW as the outcome will parallel those described above for PTD, but logistic regression techniques will be used rather than Cox regression.

Aim 2. Head-to-head comparison of treatment regimens among Cohorts A, B & C, respectively

We will obtain point and interval estimates of relative hazards of PTD associated with the three pairwise contrasts among the three treatment groups. These estimates, and associated tests of significance, will be obtained using the Cox proportional hazards regression model developed in addressing Aim 1, fully adjusted for potential confounders including depression severity, as described above, via transformed linear contrasts of treatment group associated regression parameters. Based on the results of Aim 1 analyses of dose-response, we will obtain confounder adjusted relative hazards associated with PTD with levels of dose of a given treatment modality

(e.g. antidepressants, <30 days, 30-120 days, > 120 days duration during pregnancy), vs. particular levels of dose of another treatment (e.g. psychotherapy, ≥4 visits that has shown to be effective). We would categorize the comparator treatment (e.g., psychotherapy), based on the Aim 1 findings that the effectiveness of the comparator treatment in reducing risk of PTD varied by dose (e.g. no effect for < 4 visits).

As a secondary (exploratory) analysis, we will evaluate the treatment effectiveness separately for those with more severe underlying depression before treatment vs. those with more mild depression. Treating depression may be more beneficial for those with severe depression.

Analyses of LBW as the outcome will parallel those described above for PTD, but logistic regression techniques will be used rather than Cox regression.

Aim 3. Risk of PTD & LBW for untreated depression (Cohort D vs. E)

Cox proportional hazards regression will be used to obtain point and interval estimates of the hazard ratio for PTD associated with untreated depression, vs. no depression, with adjustment for potential confounders, as described above. Parameter estimation, the approach to inclusion of potential confounders and model diagnostics are as described above for Aim 1 analyses.

To evaluate potential dose-response relationships, we will examine the strength of the association between PTD and severity of depressive symptoms. Severity will be based on results of the KPNC universal screening and clinical diagnosis of depression from the KPNC EMR. Those with a clinical diagnosis of depression will be considered to be the most severe cases followed by those who screened positive with PHQ-9 higher scores (e.g., > 20). Those who screened positive, but with lower PHQ-9 scores (10-19) will be considered to have moderate depression. The effect of severity will also be examined in terms of duration.

Analyses of LBW as the outcome will parallel those described above for PTD, but logistic regression techniques will be used rather than the Cox model.

Aim 4. Difference in treatment effectiveness by depression type (with and without comorbidity)

Analyses of Aims 1, 2 and 3, will be initially stratified by depression type, with or without other psychiatric comorbidity (yes/no). Heterogeneity in treatment effectiveness by depression type will be assessed by inclusion of appropriate cross-product (interaction) terms in regression models outlined above, using a likelihood ratio test. If there is evidence of heterogeneity, point and interval estimates of relative hazards/odds ratios of interest by comorbidity status (yes/no), adjusted for confounders, will be obtained via linear combination of regression parameter estimates (main effect and interaction terms, as appropriate).

6.1.1 Propensity Score Analysis

In addition to the conventional analyses outlined above, we will apply a propensity score (PS) analysis method to reduce or correct for any selection bias into treatment groups in our conventional approach described above. PS analysis uses information about all measured covariates to balance unobserved factors between treatment groups. It is not necessary to know the exact relationship of unobservable factors with measured covariates - as long as there is any type of relationship between unobservable and measured factors, propensity score analysis can reduce over 90% of selection bias associated with unobserved factors [80]. Given the study aims, which primarily focus on various pair-wise comparisons of treatment groups, PS analyses will be

conducted in the cohort subsets consisting of each pair of groups of interest. This will allow us to construct a multivariable logistic model for dichotomous outcomes to calculate a propensity score, i.e., the probability of receiving a given treatment, vs. another treatment (or no treatment), as a function of all known factors that might affect the treatment outcome. This model will include the same variables used in our conventional Cox proportional hazards/logistic regression models described above. The resulting probabilities (propensity scores) will be used to analyze the extent of overlap on all covariates, using graphical methods and frequency distributions. Areas of non-overlap in score between treatment groups will be trimmed (removed) from analyses. Stratified Cox proportional hazards models, with stratification on deciles of propensity score, will be used to compare the two treatment groups with respect to PTD risk [81, 82]. Logistic regression analyses of low birthweight will include the categorized score as a covariate. We will compare results from our conventional regression models with results from the PS analysis. We note that results with adjustment for confounding via propensity score techniques are not expected to appreciably differ from results using traditional regression adjustment[81, 83]. If we find that results differ, we will report the PS analysis results as our main findings due to the correction for selection bias in our measured characteristics.

6.1.2 Instrumental Variable Analysis

We will also use instrumental variable analysis (IVA) methods to assess the sensitivity of our results to unmeasured confounders. IVA is an alternative method for reducing selection bias that addresses both the effects of unobservable characteristics and the issue of dual causality between the choice of a treatment and the health outcome[84-86]. IVA methodology uses an indirect attribute that is closely associated with the type of treatment a patient receives, recreating randomized treatment assignment from a trial. This type of analysis requires identifying at least one factor that significantly affects treatment choice but is unrelated to the health outcome. IVA can be used to measure treatment effects independent of selection bias, which helps inform broad policy decisions, but it is limited in guiding clinical decisions for specific population subgroups [87]. We will initially consider physician specific prescribing preference, as measured in the KPNC pharmacy database, as the instrument [88]. It is necessary to consider how the characteristics of patients vary with respect to the instrument used when interpreting results of an IVA [84]. Our access to a rich amount of data will permit a thorough evaluation of the validity of our candidate instruments. If our selected instrument appears to be valid, and we find that IVA yields different results compared with PS analysis, we will have confirmed the likelihood of unmeasured confounding in our PS analysis. Given the extensive control for confounding using our clinical data, and the large sample sizes, we anticipate PS analysis and IVA to yield similar results. However, if IVA yields results that vary from PS results, and our instruments are reasonably valid, it suggests the presence of unmeasured confounders in our treatment comparisons in the PS and traditional analyses. Thus, it will inform in the interpretation of findings with respect to possible impact of unmeasured confounders.

6.2 Avoidance of bias

The most important strength of the proposed study is the existence of the unique KPNC region-wide universal peripartum depression screening program. Such a universal screening provides a rare opportunity to conduct a large scale (N= 88,000) population-based comparative effectiveness study of depression treatment on reducing fetal risk, while avoiding potential selection bias. Without such a screening program, a large percentage of pregnant women with depression (>50%) would not have been identified, since depression during pregnancy is

substantially under-recognized and under-diagnosed [5, 14, 45]. Thus, relying solely on clinical diagnosis of depression would likely lead to self-selection bias.

Given that randomized clinical trials are not feasible for evaluating treatment effectiveness on reducing fetal risk due to ethical concerns about assigning women into a treatment group with potential fetal risk, the proposed two-stage prospective cohort study is likely one of the best alternatives to examine the comparative effectiveness of treating depression during pregnancy. In addition to significantly reducing self-selection bias, the results of universal screening will provide information on depression severity measured by PHQ-9 score for participants. This valuable information will allow us to control for not only depression (yes/no), but also depression severity (a continuous scale) at baseline. While the distribution of depression severity was comparable among treatment groups (Table 2), ability to control for confounding by indication in such a refined detail will substantially increase the validity of our findings. Such a refined control for baseline depression severity is not likely feasible in other settings without universal peripartum depression screening.

We have also implemented several analytic methods to address the issue of possible selection into various treatment groups including *propensity score* analysis for balancing unobserved factors among treatment group and *instrument variable* analysis for possible unmeasured confounders (see Analytic Methods).

Finally, the findings from this observational study based on real world clinical experience is more relevant and applicable to actual clinical practice compared to findings from much controlled settings with selected populations (RCTs).

6.2.1 Study population

The proposed study will include all pregnant KPNC members (a population-based study), thus their racial/ethnic distribution will be similar as shown in Table 4 and will closely represent the underlying population (including Medicaid/MediCal) in the service region of Northern California. The study cohorts are described above (see C.b.2. Choice of comparators).

6.2.2 Sample size and statistical power

We expect to measure primary exposures (depression and treatment), and outcomes of interest (PTD, LBW) in approximately 88,000 women (stage-one sample). Randomly sampling from the stage-one cohort, we expect to recruit 400 pregnant women from each cohort (see Figure 2). Given the lack of methods for power calculations for the semiparametric maximum likelihood estimator (SMLE) in our two-stage sampling scheme, we very conservatively present minimum detectable effect calculations assuming the second stage sample of 3,600, noting that the gains in precision and statistical power in utilizing the measurements in stage-one sample of 88,000 women will be significant. In a small simulation study described below, we examine efficiency gains of an ad-hoc two-stage relative hazard estimator currently under investigation, which is an alternative to the semiparametric maximum likelihood estimator, attractive in its simplicity. These results provide information on the efficiency gains that can be expected in our two-stage design with the maximum likelihood estimation; the maximum likelihood estimator will be more efficient than the ad-hoc. Given results of the preliminary simulation study, we also present below estimates of minimum detectable effects using the two-stage SMLE for most estimates (in brackets, assumptions described below). Power calculations are based on the likelihood ratio test in the context of a Cox proportional hazards regression analysis and a logistic regression analyses, as appropriate to the outcome [89 90].

Given previous experience [3, 91-93], we expect negligible loss to follow-up (< 1%). Based on our experience with our study population [94], we expect a rate of PTD (gestational age < 37 weeks) ranging from 8% - 11% and LBW ranging from 5% to 10%.

Each primary contrast of interest in analyses addressing **Aims 1 through 3** can be characterized as a comparison of two groups of 400 women with respect to risk of study outcomes of interest (N=800). Minimum detectable hazard ratios for PTD range from 1.69 – 1.83 across the range in expected rate of PTD (Table 5). Minimum detectable odds ratios for LBW range from 1.67 to 2.05 across the range in expected event rate.

Table 4. Comparison of Maternal Age and Race/Ethnicity for Newborns between KPNC and the General Population in the KPNC Service Regions

Characteristics	KPNC Births (%)	All births in the KPNC Service Regions (%)
Maternal age		
< 20	12	10
20-29	43	52
30+	45	38
Maternal race/ethnicity		
Asian	15	16
Black	10	10

Relevant to **Aim 4** analyses of heterogeneity in the association between treatment/depression and PTD by psychiatric comorbidity status (yes/no with 50% of total stage 2 sample in each of the two groups), the minimum detectable hazard ratio for effect modification (interaction) was calculated, assuming a pairwise comparison of treatment effect (e.g. A vs. D) and the midpoint of the range in expected PTD rate (9.5%) [N for analyses = 1600]. The hazard ratio for effect modification is interpreted as the factor by which exposure to the effect modifier (comorbidity status = yes) increases the hazard ratio associated with treatment, over and above that in those unexposed to the effect modifier (comorbidity status = no). Using methods as outlined above, the minimum detectable hazard ratio for effect modification is 1.50, and similarly for analyses of LBW, the minimum detectable odds ratio for effect modification is 1.62 (two-sided test, α=.05, power =.80). While this stratification analysis is an *a priori* aim of the study with relevant clinical implications, we will be conservative in interpretation of our findings due to increased number of comparisons from stratification by depression type, though the need to adjust for multiple comparisons remains controversial [95-97].

Table 5. Minimum detectable hazard ratios (HR) for preterm delivery and minimum detectable odds ratios (OR) for low birthweight in a pair-wise comparison of depression exposure/treatment categories (N=800); two-sided test, significance level (α) = .05, power (1-β) =.80.

Preterm Delivery (HR)		Low Birthweight	
Rate = 8%	Rate = 11%	Rate = 5%	Rate = 10%

<u>Efficiency increase resulting from two-stage study design:</u> A preliminary simulation study has been conducted providing information on the expected efficiency gains of the two-stage SMLE, relative to an analysis of the stage-two subsample only. For our preliminary simulation study, we made a few simplifying assumptions. We assumed three depression status categories, none, depressed without treatment, depression with treatments, with population prevalence set at 70%, 25% and 5%, and outcome relative risks of 1.0 (ref), 2.0 and 3.0, respectively. We assumed one binary confounder with varying prevalence across depression categories and strength of association with outcome. The event incidence was assumed at 5% (among those unexposed to depression and confounder). The stage-one sample size was 88,000, with random selection of 400 in each of the three exposure categories. Across the scenarios considered relative efficiencies [i.e. Var(stage 2

MLE) / Var(SMLE)] ranged from 4 to 82, demonstrating substantial gains by utilizing exposure and outcome information in the full sample of 88,000.

As a very crude approximation to the impact of incorporating the exposure and outcome information on 88,000 women, we took an "effective sample size" approach, and assumed the worst case among the simulation scenarios considered, where the relative efficiency of the SMLE was approximately 4.0. Thus, rather than basing power calculations on the stage-two sample size of 800, we based them on 3200 (i.e. 4.0 * 800; 1,600 in each cohort category). Under these assumptions, the minimum detectable PTD hazard ratios are significantly decreased, ranging from 1.32 to 1.38 (for PTD rate ranging from 8%-11%) in pair-wise comparisons of depression exposure/treatment categories of interest. Similarly, the minimum detectable LBW odds ratios range from 1.32 to 1.45 (for LBW rate ranging from 5% to 10%). In addition, taking the same approach to estimation of efficiency gains, the minimum detectable hazard ratio and odds ratio for effect modification by comorbidity status (yes vs. no) are 1.31 and 1.37 in analyses of PTD and LBW, respectively. We acknowledge that these estimates of gains in efficiency are based on results from a simulation study which did not reflect the complexities in analyses that this study will encounter (e.g. multiple confounders, both multicategory and continuous).

7 STUDY ADMINISTRATION

7.1 Data Collection and Management

7.1.1 Data sources

KPNC EMRs and other clinical databases will be used to identify all women who meet our recruitment criteria and to include them in the study. Approximately 88,000 pregnant KPNC members who are 18 years or older, will be identified and recruited to the study. All eligible pregnant women will be classified into one of the categories based on the diagnosis and treatment of their depression during pregnancy. For each of these nine categories, 400 women (for a total of 3,600) will be randomly selected and recruited to conduct telephone interviews to collect information on confounders and effect modifiers.

Birth outcomes for all participants will be ascertained through the KPNC EMR and other clinical databases

7.2 Confidentiality

All standard procedures following HIPAA requirements will be implemented for this study. Names and other identifying information on study subjects will be obtained for record-keeping purposes only, and no individuals will be identified in any reports from this study. Only persons directly involved in the study will have access to data identifying individual subjects. Records and forms with identifying information will be kept in locked drawers when not in use. Access to computerized information will require simultaneous knowledge of the database, language, file names, and multiple passwords.

7.3 Risks and Benefits

7.3.1 Potential and anticipated risks to research participants.

For participants in the telephone interview (stage-two study), the risk of participating in this study is minimal. Women who had adverse pregnancy outcomes in the past may become

upset when recalling these events. All interviewers will be trained on how to appropriately handle these situations in case they arise. Our experience also indicates that some women find it therapeutic to discuss these events.

7.3.2 Risk Minimization

No invasive procedures are involved in the study protocol, only interviews. All interviewers have had experience interviewing pregnant women and are specifically trained for these types of studies. All of the women contacted for a telephone interview will have received a letter informing them about the study.

7.3.3 Benefits to participants and to society

No direct benefit is expected to participants resulting from participation in the study. Findings from the study in a publishable format will be made available to all participants who request them. The potential benefit to society is to enhance the understanding of the risk-benefit profiles of various treatment options depression during pregnancy to prevent preterm delivery and low birthweight. Because of the prevalence of depresion among pregnant women, findings from the proposed study will have a significant public health impact. Correct treatment choices may lead to reduction of PTD and LBW which remain top public health challenges globally.

7.3.4 How potential risks are justified by potential benefits

There is only minimal risk to participants. Potential benefits to society described above can be significant.

7.4 Recruitment Strategy

The chiefs and other obstetricians in the departments of Obstetrics and Gynecology at participating KPNC facilities will be informed of the nature of the study and permission to contact their patients (if selected) in stage-two will be obtained from those departments. Informational posters will be sent to all OBGYN departments for display and OBGYNS will inform all pregnant women about our study during their prenatal visits. For stage-two samples, after they are identified, a letter explaining the study and requesting their participation will be sent to women. A postage-paid and self-addressed refusal card will be included in the letter for participants who choose to refuse. For those who do not return refusal postcards, interviewers will contact women to answer their questions and ascertain their willingness to participate in the study.

7.5 Informed Consent and HIPAA Authorization

7.5.1 Waiver of Consent for Stage-One Sample

A waiver of consent has been granted for the stage-one sample, since the stage-one sample involves data only, there are no risks associated with recruiting subjects, although there is always a small chance of the loss of confidentiality during data analyses. Since there is no direct contact of patients, we do not anticipate an adverse effect on their rights or welfare.

7.5.2 Waiver of Consent for Stage-Two Sample

A waiver of the requirement for participants to sign a written consent form has been granted for the stage-two telephone interviews. As has been used in other studies of telephone interviews, for those who will be contacted by telephone for an interview, the consent process will be conducted over the phone. Verbal Informed consent will be obtained before each interview and no interviews will be conducted without consent for participation. Interviewers who are employees of Division of Research and who have had extensive experience conducting such interviews will obtain the consent. There is no more than minimal risk to participants and there are no other procedures involved beyond an interview.

7.5.3 Waiver of HIPAA Authorization

A waiver of HIPAA Authorization has been granted for both stages of the study.
The stage-one sample will be data only. There will be no direct contact with participants. For the stage-one sample of 88,000 women, we will need to access PHI to identify maternal pregnancy and depression status. The medical record number will be used to link the mother-infant pairs and other datasets on pregnancy, maternal, and birth outcomes. Stage-two will involve telephone interviews for which we are requesting a waiver of written consent for participation (verbal consent will be obtained).

Data with identifiable information will be password protected by the programmer. An additional dataset will be created and used by the programmer which will not include any PHI. Data will not be shared outside of KPNC.

No PHI data will be disclosed to outside investigators. We will provide assurance of the confidentiality and proper procedures for protecting PHI as set forth by the KPNC IRB.

Personal identifiers will be destroyed at the end of study according to IRB requirements.

7.5.4 Payment to Subjects

After the interview, a thank-you letter with a check for $40 will be sent to each participant for their time and out-of-pocket expenses.

8 DISSEMINATION

A Stakeholder Advisory Board (SAB) will be closely involved in the dissemination of study findings and implementation assessment. The Stakeholder Advisory Board (SAB) will be established to ensure that the study addresses the concerns of pregnant women and providers about prenatal depression and its treatment (see Patient and Stakeholder Engagement Plan for detail). This board will be comprised of six stakeholders: two patient partners, two women who were diagnosed with depression during pregnancy, and two health care providers. The two patient organizations (see below) that have enthusiastically supported the proposed study, were identified as appropriate partners for this study because, among other things, they have many years of first-hand experience disseminating scientific information in the areas of pregnancy and mental health to their members nationwide including providers and pregnant women.

The SAB members are valuable facilitators to dissemination and implementation of study findings. Our collaboration with *Childbirth Connection* and *CMMHC* will enable us to disseminate study findings to a broader audience through the use of their extensive *statewide and national*

networks. Their constituents will be informed of our study findings. For dissemination within KPNC, Drs. Flanagan and Turner together with co-investigators, Drs. Young and Hamilton, working with the OB/GYNs and mental health care providers at KPNC will greatly facilitate both the dissemination and implementation of the study findings as well as incorporation of the study findings into practice guidelines for treating depression among pregnant women. They will also be champions for disseminating study findings through KPNC existing member communication channels.

8.1 Dissemination to Patients and Advocates (statewide and nationwide)

The California Maternal Mental Health Collaborative (CMMHC) consists of medical and mental health professionals (i.e., providers), educators, community advocates, and individuals who have themselves experienced mental health disorders (i.e., patients) and their mission is to increase and improve maternal mental health awareness, diagnosis and treatment. CMMHC utilize educational pamphlets, their website and social networking media outlets such as Facebook, twitter and blogs to disseminate study findings to their members and communities throughout the nation. They will disseminate our study findings throughout their nationwide networks.

Childbirth Connection has a history of providing national leadership in identifying, demonstrating and fostering innovative ways to improve maternity care for all women and their families using research, education and advocacy. Childbirth Connection will disseminate the findings from our study through their online evidence-based maternity care resource directory, among other methods nationwide. This resource directory provides a way to disseminate evidence based medicine information to those who want to plan, practice, and receive care within this essential framework. Our findings will also be distributed through their Transforming Maternity Care Blog which features news, opinions and analyses from Childbirth Connection and prominent guest contributors. Other dissemination tools will include reports, fact sheets, pamphlets and other handouts which will also be used to disseminate findings from the study to pregnant women and women of child-bearing age.

8.2 Dissemination to Providers within KPNC

Drs. Flanagan (OBGYN) and Mason (Mental & Behavioral Health), as leaders in their respective specialties at KPNC, will work with KPNC leadership to implement the findings from this study into practice throughout the Kaiser Permanente Northern California Region. They will also work with health care providers and educators to disseminate the findings among both providers and patients, and encourage providing depression treatment among pregnant women based on findings from this study.

8.3 Making study results available to study participants after completion of analyses.

We will work with Childbirth Connection and CMMHC to develop a brief summary of the results from our study for the general public, including our participants. Both of these organizations have a long-history of developing educational materials for patients and the general public and we will use their expertise to guide development of the materials for dissemination to our study participants. The summary will then be sent to our participants according to their preference (US postal service, email, link to organizational websites, etc.)

9 REFERENCES

1. 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. Obstet Gynecol 2009;114(3):703-713.
2. Gaynes BN, Gavin N, Meltzer-Brody S et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evid Rep Technol Assess (Summ) 2005;(119):1-8.
3. Li D, Liu L, Odouli R. Presence of depressive symptoms during early pregnancy and the risk of preterm delivery: a prospective cohort study. Hum Reprod 2009;24(1):146-153.
4. Dietz PM, Williams SB, Callaghan WM, Bachman DJ, Whitlock EP, Hornbrook MC. Clinically identified maternal depression before, during, and after pregnancies ending in live births. Am J Psychiatry 2007;164(10):1515-1520.
5. Mann JJ. The medical management of depression. N Engl J Med 2005;353(17):1819-1834.
6. Evans J, Heron J, Francomb H, Oke S, Golding J. Cohort study of depressed mood during pregnancy and after childbirth. BMJ 2001 Aug 4 ;323 (7307):257 -60323(7307):257-260.
7. Chung TK, Lau TK, Yip AS, Chiu HF, Lee DT. Antepartum depressive symptomatology is associated with adverse obstetric and neonatal outcomes. Psychosom Med 2001 Sep -Oct ;63 (5):830 -463(5):830-834.
8. Fergusson DM, Horwood LJ, Thorpe K. Changes in depression during and following pregnancy. ALSPAC Study Team. Study of Pregnancy and Children. Paediatr Perinat Epidemiol 1996;10(3):279-293.
9. Homish GG, Cornelius JR, Day NL, Goldschmidt L. Depressive symptomatology in perinatal women. American Journal of Epidemiology 155, S26. 2002.
10. Coates AO, Schaefer CA, Alexander JL. Detection of postpartum depression and anxiety in a large health plan. J Behav Health Serv Res 2004;31(2):117-133.
11. Perez-Stable EJ, Miranda J, Munoz RF, Ying YW. Depression in medical outpatients. Underrecognition and misdiagnosis. Arch Intern Med 1990;150(5):1083-1088.
12. Chalem E, Mitsuhiro SS, Manzolli P et al. Underdetection of psychiatric disorders during prenatal care: a survey of adolescents in Sao Paulo, Brazil. J Adolesc Health 2012;50(1):93-96.
13. Rowan P, Greisinger A, Brehm B, Smith F, McReynolds E. Outcomes from implementing systematic antepartum depression screening in obstetrics. Arch Womens Ment Health 2012;15(2):115-120.
14. Josefsson A, Berg G, Nordin C, Sydsjo G. Prevalence of depressive symptoms in late pregnancy and postpartum. Acta Obstet Gynecol Scand 2001;80(3):251-255.
15. Marcus SM, Flynn HA, Blow FC, Barry KL. Depressive symptoms among pregnant women screened in obstetrics settings. J Womens Health (Larchmt) 2003;12(4):373-380.
16. Cooper PJ, Landman M, Tomlinson M, Molteno C, Swartz L, Murray L. Impact of a mother-infant intervention in an indigent peri-urban South African context: pilot study. Br J Psychiatry 2002 Jan ;180 :76 -81180:76-81.
17. Beck S, Wojdyla D, Say L et al. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. Bull World Health Organ 2010;88(1):31-38.
18. Bettegowda VR, ackritz E, etrini JR. Chapter 3: Epidemiologic Trends in Perinatal Data. March of Dimes Foundation, 2011

19. Simhan HN, Caritis SN. Prevention of preterm delivery. N Engl J Med 2007;357(5):477-487.
20. Berkowitz GS, Papiernik E. Epidemiology of preterm birth. Epidemiol Rev 1993;15:414-443.
21. Kline J, Stein Z, Susser M. Conception to Birth: Epidemiology of Prenatal Development. 1st ed. Oxford University Press Inc.; 1989.
22. Goldenberg RL, Rouse DJ. Prevention of premature birth. N Engl J Med 1998;339:313-320.
23. Board of Health Sciences Policy. Societal Costs of Preterm Birth. In: Behrman R, Butler A, editors. Preterm Birth: Causes, Consequences, and Prevention. 2007:398.
24. Cuevas KD, Silver DR, Brooten D, Youngblut JM, Bobo CM. The cost of prematurity: hospital charges at birth and frequency of rehospitalizations and acute care visits over the first year of life: a comparison by gestational age and birth weight. Am J Nurs 2005;105(7):56-64.
25. Armstrong J. 17 Progesterone for preterm birth prevention: a potential 2 billion dollar opportunity. Am J Obstet Gynecol 2007;196(3):194-195.
26. MARCH OF DIMES. WHITE PAPER ON PRETERM BIRTH: THE GLOBAL AND REGIONAL TOLL. 2009
27. Beck S WDSLBAMMRJRCMRVLP. WHO systematic review on maternal mortality and morbidity: The global burden of preterm birth. 2009
28. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet 2008;371(9606):75-84.
29. Kramer MS, Platt R, Yang H et al. Secular trends in preterm birth: a hospital-based cohort study. JAMA 1998;280(21):1849-1854.
30. Branum AM, Schoendorf KC. Changing patterns of low birthweight and preterm birth in the United States, 1981-98. Paediatr Perinat Epidemiol 2002;16(1):8-15.
31. Alder J, Fink N, Bitzer J, Hosli I, Holzgreve W. Depression and anxiety during pregnancy: a risk factor for obstetric, fetal and neonatal outcome? A critical review of the literature. J Matern Fetal Neonatal Med 2007;20(3):189-209.
32. Mastorakos G, Pavlatou MG, Mizamtsidi M. The hypothalamic-pituitary-adrenal and the hypothalamic- pituitary-gonadal axes interplay. Pediatr Endocrinol Rev 2006;3 Suppl 1:172-181.
33. Chrousos GP, Torpy DJ, Gold PW. Interactions between the hypothalamic-pituitary-adrenal axis and the female reproductive system: clinical implications. Ann Intern Med 1998;129(3):229-240.
34. Mastorakos G, Ilias I. Maternal hypothalamic-pituitary-adrenal axis in pregnancy and the postpartum period. Postpartum-related disorders. Ann N Y Acad Sci 2000;900:95-106.
35. Mastorakos G, Ilias I. Maternal and fetal hypothalamic-pituitary-adrenal axes during pregnancy and postpartum. Ann N Y Acad Sci 2003;997:136-149.
36. Orr ST, James SA, Blackmore PC. Maternal prenatal depressive symptoms and spontaneous preterm births among African-American women in Baltimore, Maryland. Am J Epidemiol 2002;156(9):797-802.
37. Dayan J, Creveuil C, Herlicoviez M et al. Role of anxiety and depression in the onset of spontaneous preterm labor. Am J Epidemiol 2002;155(4):293-301.
38. Steer RA, Scholl TO, Hediger ML, Fischer RL. Self-reported depression and negative pregnancy outcomes. J Clin Epidemiol 1992;45(10):1093-1099.
39. Dole N, Savitz DA, Siega-Riz AM, Hertz-Picciotto I, McMahon MJ, Buekens P. Psychosocial factors and preterm birth among African American and White women in central North Carolina. Am J Public Health 2004;94(8):1358-1365.

40. Lumley J, Austin MP. What interventions may reduce postpartum depression. Curr Opin Obstet Gynecol 2001 Dec ;13 (6):605 -1113(6):605-611.

41. Hann D, Winter K, Jacobsen P. Measurement of depressive symptoms in cancer patients: evaluation of the Center for Epidemiological Studies Depression Scale (CES-D). J Psychosom Res 1999;46(5):437-443.

42. Roberts RE. Reliability of the CES-D Scale in different ethnic contexts. Psychiatry Res 1980;2(2):125-134.

43. Roberts RE, Vernon SW. The Center for Epidemiologic Studies Depression Scale: its use in a community sample. Am J Psychiatry 1983;140(1):41-46.

44. Myers JK, Weissman MM. Use of a self-report symptom scale to detect depression in a community sample. Am J Psychiatry 1980;137(9):1081-1084.

45. Evans J, Heron J, Patel RR, Wiles N. Depressive symptoms during pregnancy and low birth weight at term: longitudinal study. Br J Psychiatry 2007;191:84-5.:84-85.

46. Diego MA, Field T, Hernandez-Reif M, Schanberg S, Kuhn C, Gonzalez-Quintero VH. Prenatal depression restricts fetal growth. Early Hum Dev 2009;85(1):65-70.

47. Maina G, Saracco P, Giolito MR, Danelon D, Bogetto F, Todros T. Impact of maternal psychological distress on fetal weight, prematurity and intrauterine growth retardation. J Affect Disord 2008;111(2-3):214-220.

48. Chambers CD, Johnson KA, Dick LM, Felix RJ, Jones KL. Birth outcomes in pregnant women taking fluoxetine. N Engl J Med 1996;335(14):1010-1015.

49. Chambers CD, Hernandez-Diaz S, Van Marter LJ et al. Selective serotonin-reuptake inhibitors and risk of persistent pulmonary hypertension of the newborn. N Engl J Med 2006;354(6):579-587.

50. Tuccori M, Testi A, Antonioli L et al. Safety concerns associated with the use of serotonin reuptake inhibitors and other serotonergic/noradrenergic antidepressants during pregnancy: a review. Clin Ther 2009;31 Pt 1:1426-53.:1426-1453.

51. Moulsdale W, Hermann S. In Utero Exposure to Selective Serotonin Reuptake Inhibitors: Evidence for Poor Neonatal Adaptation. Newborn and Infant Nursing Reviews 2008.

52. Dimidjian S, Goodman S. Nonpharmacologic intervention and prevention strategies for depression during pregnancy and the postpartum. Clin Obstet Gynecol 2009;52(3):498-515.

53. 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.

54. Cohen LS, Altshuler LL, Harlow BL et al. Relapse of major depression during pregnancy in women who maintain or discontinue antidepressant treatment. JAMA 2006;295(5):499-507.

55. Toh S, Mitchell AA, Louik C, Werler MM, Chambers CD, Hernandez-Diaz S. Antidepressant use during pregnancy and the risk of preterm delivery and fetal growth restriction. J Clin Psychopharmacol 2009;29(6):555-560.

56. Davis RL, Rubanowice D, McPhillips H et al. Risks of congenital malformations and perinatal events among infants exposed to antidepressant medications during pregnancy. Pharmacoepidemiol Drug Saf 2007;.

57. Djulus J, Koren G, Einarson TR et al. Exposure to mirtazapine during pregnancy: a prospective, comparative study of birth outcomes. J Clin Psychiatry 2006;67(8):1280-1284.

58. Chambers CD, Johnson KA, Dick LM, Felix RJ, Jones KL. Birth outcomes in pregnant women taking fluoxetine. N Engl J Med 1996;335(14):1010-1015.

59. Ericson A, Kallen B, Wiholm B. Delivery outcome after the use of antidepressants in early pregnancy. Eur J Clin Pharmacol 1999;55(7):503-508.

60. Simon GE, Cunningham ML, Davis RL. Outcomes of prenatal antidepressant exposure. Am J Psychiatry 2002;159(12):2055-2061.

61. Wen SW, Yang Q, Garner P et al. Selective serotonin reuptake inhibitors and adverse pregnancy outcomes. Am J Obstet Gynecol 2006;194(4):961-966.

62. Yonkers KA, Norwitz ER, Smith MV et al. Depression and Serotonin Reuptake Inhibitor Treatment as Risk Factors for Preterm Birth. Epidemiology 2012.

63. Kulin NA, Pastuszak A, Sage SR et al. Pregnancy outcome following maternal use of the new selective serotonin reuptake inhibitors: a prospective controlled multicenter study. JAMA 1998;279(8):609-610.

64. Dayan J, Creveuil C, Marks MN et al. Prenatal depression, prenatal anxiety, and spontaneous preterm birth: a prospective cohort study among women with early and regular care. Psychosom Med 2006;68(6):938-946.

65. Ostergaard SD, Waltoft BL, Mortensen PB, Mors O. Environmental and familial risk factors for psychotic and non-psychotic severe depression. J Affect Disord 2013;147(1-3):232-240.

66. Nelson EB. Psychotic depression--beyond the antidepressant/antipsychotic combination. Curr Psychiatry Rep 2012;14(6):619-623.

67. Farahani A, Correll CU. Are antipsychotics or antidepressants needed for psychotic depression? A systematic review and meta-analysis of trials comparing antidepressant or antipsychotic monotherapy with combination treatment. J Clin Psychiatry 2012;73(4):486-496.

68. Wijkstra J, Lijmer J, Balk F, Geddes J, Nolen WA. Pharmacological treatment for psychotic depression. Cochrane Database Syst Rev 2005;(4):CD004044.

69. Wijkstra J, Schubart CD, Nolen WA. Treatment of unipolar psychotic depression: the use of evidence in practice guidelines. World J Biol Psychiatry 2009;10(4 Pt 2):409-415.

70. Pearlstein T. Perinatal depression: treatment options and dilemmas. J Psychiatry Neurosci 2008;33(4):302-318.

71. Dimidjian S, O'Hara MW. Pharmacotherapy or untreated antenatal depression: a false dichotomy. J Clin Psychiatry 2009;70(9):1321-1322.

72. Boath E, Bradley E, Henshaw C. The prevention of postnatal depression: a narrative systematic review. J Psychosom Obstet Gynaecol 2005;26(3):185-192.

73. Misri S, Kendrick K. Treatment of perinatal mood and anxiety disorders: a review. Can J Psychiatry 2007;52(8):489-498.

74. ACOG. Ob-Gyns Encouraged to Screen Women for Depression During and After Pregnancy . 2010

75. 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. Arch Gen Psychiatry 2010;67(10):1012-1024.

76. Anderson PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat 1982;10:1100-1120.

77. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer-Verlag, 2000

78. Scott AJaWCJ. Calculating efficient semiparametric estimators for a broad class of missing-data problems. Dept. of Mathematics, Statistics and Philosophy, Univ. of Tampere; 2006.

79. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. Am J Epidemiol 1989;129(1):125-137.

80. Rosenbaum PR, Rubin DB. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. Journal of the American Statistical Association 1984;79(387):516-524.

81. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol 2006;59(5):437-447.

82. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol 2006;98(3):253-259.

83. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. J Clin Epidemiol 2005;58(6):550-559.

84. Brooks JM, Chrischilles EA. Heterogeneity and the interpretation of treatment effect estimates from risk adjustment and instrumental variable methods. Med Care 2007;45(10 Supl 2):S123-S130.

85. Earle CC, Tsai JS, Gelber RD, Weinstein MC, Neumann PJ, Weeks JC. Effectiveness of chemotherapy for advanced lung cancer in the elderly: instrumental variable and propensity analysis. J Clin Oncol 2001;19(4):1064-1070.

86. Hadley J, Polsky D, Mandelblatt JS et al. An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population. Health Econ 2003;12(3):171-186.

87. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA 2007;297(3):278-285.

88. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. Int J Biostat 2007;3(1):14.

89. Self SG, Mauritsen RJ, Ohara J. Power calculations for likelihood ratio tests in generalized linear models. Biometrics 1992;48:31-39.

90. Egret Siz Reference Manual. Manual Revision 10. Seattle: Statistics and Epidemiology Research Corporation; 1992.

91. Li DK, Odouli R, Wi S et al. A population-based prospective cohort study of personal exposure to magnetic fields during pregnancy and the risk of miscarriage. Epidemiology 2002;13(1):9-20.

92. Li DK, Liu L, Odouli R. Exposure to non-steroidal anti-inflammatory drugs during pregnancy and risk of miscarriage: population based cohort study. BMJ 2003;327(7411):368.

93. Li DK, Janevic T, Odouli R, Liu L. Hot tub use during pregnancy and the risk of miscarriage. Am J Epidemiol 2003;158(10):931-937.

94. Li DK, Odouli R, Liu L, Vinson M, Trachtenberg E. Transmission of parentally shared human leukocyte antigen alleles and the risk of preterm delivery. Obstet Gynecol 2004;104(3):594-600.

95. Rothman K.J. No adjustments are needed for multiple comparisons. Epidemiology. 1990:43-46.

96. Savitz DA. Interpreting Epidemiologic Evidenc: Strategies for Study Design and Analysis. New York: Oxford University Press; 2003.

97. Rothman K.J. Modern Epidemiology. Boston: Little, Brown & Co; 1998.