Original Trial Protocol

Date of Application: 2010-05-13

Title of Study: Depression, Estrogen Replacement, and Cardiovascular Health in the Perimenopause

Name and degrees of Principal Investigator(s): Susan S. Girdler, PhD & David Rubinow, M.D.

Name of funding source or sponsor: National Institute of Mental Health

<u>Amendments Made to the Following Protocol:</u>

Unless otherwise stated, the following changes were made to the study exclusion criteria to facilitate participant recruitment:

- Women with borderline elevated LDL (190 – 200 mg/dl) at baseline were allowed to enter the trial if their physician was not intending to treat them over the course of the subsequent 12 months **(2010-11-16).**
- Early postmenopausal women (in STRAW +1b) were allowed to enter the trial **(2010-12-30).**
- Women with a history of >1 biopsy were allowed to enter the trial if there was no personal or family history of breast cancer in the first or second degree relatives and if the documented histologic findings for all biopsies were consistent with the following: fibrocystic changes (within the normal range - cysts and ductal ectasia, mild hyperplasis, nonsclerosing adenosis, and periductal fibrosis, simple fibroadenoma, and miscellaneous (lobular hyperplasia, juvenile hypertrophy, and stromal hyperplasia)); benign tumors (hamartoma, lipoma, phyllodes tumor, solitary papilloma, neurofibroma, giant adenoma, and adenomyoepithelima); traumatic lesions (hematoma, fat necrosis, and lesions caused by penetration by a foreign body); infections (granuloma and mastitis); sarcoidosis; metaplasia (squamous and apocrine); diabetic mastopathy **(2010-12-30).**
- The requirement that all women have an FSH level >2 SD above mean premenopausal levels was removed **(2011-02-17)**.
- Women without a uterus but with at least one ovary retained were allowed to enter the trial if: 1) they were experiencing vasomotor symptoms and their baseline estradiol levels were above postmenopausal concentrations (> 40 pg/ml) or, 2) they were not experiencing vasomotor symptoms but had baseline estradiol > 40 pg/ml and baseline FSH levels > 14 pg/ml, which is two standard deviations above the mean level of a previously-recruited sample of reproductive-aged women, in line with STRAW staging guidelines **(2011-03-04).**
- If a participant knew the complete family history for the mother's side but none or only partial history for the father's side, it was decided that the study gynecologist would evaluate whether transdermal estradiol would be contraindicated based on the history provided – if not, she could enter the trial **(2012-09-20).**
- Women with a single episode of past depression were considered eligible if their single episode met the following criteria: the duration was at least 3 months, at least 6 of 9 MDD

symptoms were experienced (exceeding the required 5-symptom threshold), and the episode occurred as an adult (not in childhood or early adolescence) **(2013-08-05).**

- The acceptable age range extended from 45-55 to 45-60 **(2014-02-27)**.
- The exclusion criteria related to psychiatric history were changed to exclude women with: history of suicide attempts; history of bipolar and related disorders or schizophrenia spectrum and other psychotic disorders; history of severe substance use if it was in the past 10 years; less than two years in remission from substance use disorders, bulimia nervosa, anorexia nervosa, or posttraumatic stress disorder; current psychiatric diagnosis of binge eating disorder, a neurodevelopmental disorder, or a trauma- and stress-related disorder (except for posttraumatic stress disorder) with severity greater than mild **(2014-02-27).**
- Women having had a bilateral oophorectomy were allowed into the trial if the procedure had occurred within 24 months and the participant had been menstruating regularly prior to the procedure **(2014-02-27).**
- The requirement of having a baseline CES-D score ≤8 was increased to ≤16 **(2010-11-16)** and later removed altogether **(2014-02-27).**
- The acceptable body mass index range was lowered to <30 due to concerns that estradiol therapy may be less safe in obese women **(2011-08-18)** but was later increased to <35 based on research suggesting that transdermal estradiol, specifically, was not related to an increased health risk in these women **(2014-04-04)**.

Changes made to study protocol:

- Rather than starting a participant on 0.1 mg of estradiol immediately, the protocol was changed so that participants started on 0.025 mg for two weeks, then 0.05 mg patch for the next four weeks, and a 0.1 mg patch for the remainder of the study **(2011-02-14).**
- To avoid very heavy bleeding following progesterone administration, the schedule for administering progesterone was changed from every three months to every two months **(2011-07-08).**
- Participant compensation was increased to $1325 when study visits at months 6 and 12 began occurring separately from the months 6 and 12 laboratory visits **(2012-05-15).**
- Following randomization, participants were given the option to participate in four follow-up phone calls in which the CES-D was administered (months 15, 18, 21 and 24) for an additional $100 in compensation **(2012-05-15).**

Brief Summary

Purpose: The purpose of this study is threefold: 1) to examine cardiovascular risk factors in perimenopausal women with or without histories of recurrent depression; and the role of cumulative trauma exposure in the relationship of histories of depression to cardiovascular risk; 2) to examine progression in cardiovascular risk and the development of depressive episodes over 12 months in perimenopausal women with or without a history of recurrent depression who are not being treated with estradiol replacement therapy (ERT); and 3) to examine predictors and mediators of ERT - related treatment efficacy for both depression and cardiovascular risk.

Participants: Three hundred and twenty (320) women, 45 – 55 years of age undergoing a natural perimenopause (≥ 2 skipped cycles and an interval of amenorrhea ≥ 60 days with FSH > 2 SD above mean premenopausal levels) will be tested such that half (n= *160*) will have a history of recurrent DEP (2+ episodes of depression, at least one of which is major depression) but be in remission, while the other half will be recruited as never depressed controls, matched for age and current low depressive symptoms (≤ 8 on the CES-D 20-item depression scale). All subjects will be medically healthy perimenopausal women, 45 – 55 years of age with no contraindications for ERT.

Procedures (methods): Subjects will be tested at baseline for indices of cardiovascular risk (cardiovascular stress reactivity, metabolic risk, cardiac autonomic tone, and endothelial function) and for depressive symptoms, functional impairment, menopausal symptoms, negative life events, diet and exercise. Lifetime trauma exposure will also be determined. Using a placebo-controlled, double-blind design, following baseline assessments, half of the women in each group (history of depression or never depressed) will be randomized to either transdermal 17β-E2 (100ug/day) or placebo for *12* months (oral micronized progesterone will be given every 3 months to women on active ERT to protect the endometrium and placebo progesterone given to women on placebo ERT). Subjects will be assessed at months 1, 2, 4, 6, 8, 10 and 12 for compliance, vitals, side effects, depressive symptom severity and functional impairment (also menopausal symptoms, diet, and exercise). At months 6 and 12, all CV risk measures conducted at the baseline (described above) will be repeated.

Purpose and Rationale

The following are the driving assumptions of this study:

1. Depression and CVD display reciprocal risks: Both animal and human studies demonstrate a strong link between DEP – a major source of morbidity – and CVD – the leading killer of women in the US. Patients with CVD are more likely to develop DEP (e.g., the prevalence of DEP in patients with congestive heart failure reportedly reaches 40%);, and conversely, the presence of DEP in patients with CVD predicts increased morbidity and mortality. This risk applies as well to women with depressive symptoms (50% greater likelihood of dying a cardiac death) and even euthymic individuals with a history of DEP (see below).

2. The risk of DEP and CVD increases during the perimenopause: Although age is likely the major cause of the rapid increase in the incidence of CVD in PM women, the importance of estradiol in CV risk is supported by the substantial increase in CVD in women undergoing early menopause, oophorectony without estrogen replacement, or premature ovarian failure, a disorder associated with impaired endothelial function and increased CV mortality. Studies in cynomolgus macaques also demonstrate that impaired ovarian function is associated with extensive coronary atherosclerosis. Thus, loss of estrogen appears to contribute to the higher incidence of CVD morbidity and mortality in older women. Studies have demonstrated at least a twofold increase in depression in women during the perimenopause, and our group has shown that estrogen withdrawal precipitates depression in women on hormone replacement therapy (HRT) (combined estrogen plus progestin) with a history of perimenopausal depression.

3. Depression, CVD, and estrogen deficiency share common pathophysiologic perturbations, which collectively constitute a risk profile for CVD and depression. These include the following: 1) Immune activation and inflammation; 2) Endothelial dysfunction; 3) Cardiac Autonomic Dysregulation; and 4) Metabolic Risk. It is well established that each of these factors predicts CVD and, as well, is associated with both depression and post-menopausal status.

4. A past history of depression, and factors that lead to depression (traumatic events), predispose toward an adverse profile of these mediators of pathophysiology.

Substantial evidence suggests that euthymic individuals with histories of recurrent depression show greater rates of CVD, a pro-inflammatory state, impaired endothelial function, altered cardiac autonomic control, and elevated metabolic risk at levels equivalent to those seen in current depression. Based on this and evidence suggesting that histories of depression may predict the development of depression as well as functional impairment in the perimenopause, **we will employ historical recurrent depression as a model of vulnerability to depression.** In view of the shared physiologic perturbations seen in CVD and depression, perturbations that are likely to be amplified in the setting of estrogen deficiency, we predict that during the perimenopause, women with histories of depression will demonstrate greater likelihood of affective dysregulation (depression episodes, severity of depressive symptoms, functional impairment, and affective instability) and progression of cardiovascular risk.

Because stress **elicits a cascade of interrelated physiologic responses that negatively impact each of the "pathways" of risk described above,** exaggerated stress reactivity may play a pivotal role in the pathogenesis and comorbidity of depression and CVD. Cardiovascular reactivity to stress predicts cardiac events and the progression of atherosclerosis. Stress also activates proinflammatory cytokines in the periphery and CNS. Stress responses are also likely to plays a role in metabolic risk., since glucocorticoids induce insulin resistance and individuals at metabolic risk show alterations in both sympathetic and hypothalamic-pituitary-adrenal (HPA)-axis activity. Thus, greater cardiovascular, inflammatory and HPA axis reactivities to stress (documented in both depression and in peri- and postmenopausal women), may comprise a phenotypic profile of risk that characterizes both depression and CVD. Further, a history of trauma exposure, which increases the risk for depression, metabolic disorders, and CVD and is associated with exaggerated sympathetic and HPA-axis reactivity. Thus, we predict that lifetime trauma exposure will explain some of the association between histories of depression and heightened physiologic risk during the perimenopause.

5**. Estrogen replacement therapy (ERT) in the perimenopause may prevent the progression of the pathophysiological risk profile associated with CVD and DEP and the emergence of affective dysyregulation.**

In animals and humans, E2 is associated with beneficial effects on inflammation, endothelial dysfunction, cardiac autonomic dysregulation, and metabolic risk. Our own work and that of others has also shown that in perimenopausal and early (but not late) postmenopausal women, ERT is also effective in reducing depressive symptoms in women experiencing major or minor depression, as well as subsyndromal affective symptoms. Additionally, we have demonstrated that estradiol withdrawal precipitates depression in women with a history of perimenopausal depression but not in those lacking that history.

The beneficial effects of estrogen on CV risk markers demonstrated in controlled trials is consistent with the results from the majority of observational studies showing cardioprotective effects of ERT/HRT in postmenopausal women. Several large secondary and primary prevention trials of ERT/HRT, however, have challenged the safety and efficacy of HRT. Nonetheless, secondary subgroup analyses and one secondary analyses of surrogate endpoints from the Women's Health Initiative (WHI) study are consistent with animal studies that, together, support the "timing hypothesis;" i.e., the effect of estrogen on CVD risk depends on when therapy is started relative to the onset of ovarian failure: beneficial or neutral if started in the perimenopause and adverse if started later. Recent reports from the North American Menopause Society and the Endocrine Society confirm that the risk of CVD in the WHI was observed in older but not younger women. Also, the formulation of estrogen used in the WHI (i.e., premarin) is associated with an increased profile of risk compared with transdermal estradiol.

What is not controversial is the existence of variance in both the cardioprotective and antidepressant effects of estradiol, making it critically important to identify those for whom beneficial effects of ERT on cardiovascular and mood regulation can be predicted. Consequently, the aim of the proposed research is to first identify a profile (physiologic and phenomenologic) of risk for mood dysregulation and CVD during the perimenopause, and then identify those factors that predict the greatest benefits of ERT in the prophylaxis of the progression of CV risk and appearance of depression.

Availability of predictors of increased risk of depression or CVD during the perimenopause - or of the risks or benefits of estradiol replacement - would significantly advance the goal of individualized medicine, significantly impact public health through decreasing the risk of these prevalent and disabling disorders, and potentially advance our understanding of the pathophysiology of depression unrelated to the perimenopause.

Subjects

A total of about four hundred and fifty (450) naturally perimenopaapusal women, 45 – 55 years of age, will be enrolled into the protocol. Of those we predict will reach randomization following all medical and psychiatric assessments (n=320), exactly one-half (n=160) will be specifically recruited to have a history of recurrent depression, defined as 2+ prior episodes of depression at least one of which must be an episode of major depressive disorder, while other episodes may include minor depressive disorder or dysthymic disorder. All subjects will be medically healthy volunteers with no contraindications for ERT.

Inclusion Criteria

- Natural perimenopause (≥ 2 skipped cycles and an interval of amenorrhea ≥ 60 days with FSH > 2 SD above mean premenopausal levels)
- 45 – 55 years of age
- Normal mammogram within one year of study entry
- Normal breast, pelvic and endometrial exam
- Current low level depressive symptoms (CES-D score < 8)

- Five or more years in remission from any anxiety disorders, substance abuse, or eating disorders.

General Exclusion Criteria:

- Endometrial hyperplasia (confirmed by endometrial ultrasound and/or biopsy)
- Abnormal uterine anatomy (e.g. fibroids)
- History of thrombophlebitis or thromboembolic disorders
- History of estrogen-dependent neoplasias
- Gall bladder disease
- Liver dysfunction or other disorders for which estrogen or progesterone use is contraindicated
- History of any CVD including coronary artery disease, arteriosclerosis, heart attack, or stroke
- Atrial fibrillation, frequent premature atrial or ventricular beats, or other rhythm abnormalities
- Type I or Type II diabetes
- Known sensitivities to any ingredient in the Climara® transdermal estradiol system or Prometrium® (oral micronized progesterone)
- Stage 1 or 2 hypertension (SBP ≥ 140 mmHg or DBP ≥ 90 mmHg)
- Fasting low density lipoprotein (LDL) levels ≥ 190 mg/dl
- BMI > 37
- Use of psychotropics, antihypertensives, statins, or hormonal preparations, or frequent use of anti-inflammatory agents (> 10 times/month).
- Use of herbals indicated for menopausal symptoms (e.g. black cohosh) and/or mood (e.g. St. John's Wort). Other herbals may be permitted if cleared by the study physicians.
- History of suicide attempts
- Current Axis I psychiatric diagnosis (e.g., depression, anxiety, substance abuse/dependence)
- Less then five years in remission from any anxiety, substance abuse or eating disorder
- History of substance dependence, bipolar or psychotic disorder
- Current migraine with aura or a history of migraine with aura if less than 5 years prior to enrollment
- Cigarette smoking > 10 per day
- Any history of migraine with or without aura in current cigarette smokers

Exclusion Criteria Specifically Designed to Minimize risk with respect to breast or ovarian cancer

. We have designed the following exclusion criteria to ensure, as much as possible, that no woman enrolled in the study will have greater than 'average risk' (i.e., greater risk then seen in the population at large) for breast or ovarian cancer. Although the prevalence of BRCA1/2 mutation positive breast cancer is approximately 5-10%, and it is estimated that only 1 in 250 – 500 individuals would carry the BRCA 1/2 mutation (Schwartz GF, Hughes KS, Lynch HT,

Fabian CJ, et al., (2009). Proceedings of the International Consensus Conference on Breast Cancer Risk, Genetics and Risk Management, April, 2007. The Breast Journal, Vol 15 (1), 4-16), a thorough family history assessment will strengthen our ability to identify women who might be at 'very high risk' for breast cancer due to a high penetrance gene mutation (e.g. BRCA1/2, PTEN or Tp53). Additionally, although ovarian cancer is also associated with the BRCA 1/2 mutation, ovarian cancer is listed as a 'rare disease' by the Office of Rare Diseases at the NIH and therefore affects fewer than 200,000 persons in the U.S. population. The International Consensus Conference (Schwartz et al., 2008) concluded that a cancer family history collected through three generations should be considered sufficient to identify those at risk for a high penetrance genetic mutation. The history includes that of the index patient, as well as those of progeny, siblings, and parents, together with second degree relatives and both sets of grandparents.

Thus, to reduce as much as possible any risks associated with exposure to estradiol, we will obtain a comprehensive family history of cancer in each of our subjects to ensure that women enrolled into the present study would meet clinical criteria for 'average risk' (i.e., risk seen in the female population at large) of breast and ovarian cancer. Based on expert recommendations laid out in the report from The International Consensus Conference on Breast Cancer Risk, Genetics, and Risk Management (Schwartz et al., 2008) as well as the Gail model (http://www.cancer.gove/bcrisktool), a statistical model developed from a number of national large-scale studies (and currently the model approved by the FDA for determining who should be offered tamoxifen or raloxifene), we propose the following inclusion/exclusion criteria based on personal and family history (See attached self-report instrument that we intend to employ).

We will exclude women with:

- a personal history of breast cancer
- a personal history of more than one breast biopsy, or a personal history of even one breast biopsy with atypical hyperplasia
- more than one first degree relative with breast cancer
- premenopausal breast cancer in even one first degree relative,
- more than three first, second, or third-degree relatives with breast cancer regardless of age
- two or more first degree relatives with any cancer with onset before age 60
- multiple primary cancers in a single relative (particularly breast and ovarian)
- any male breast cancer (as it is almost exclusively related to the BRCA 1/2 mutation)
- ovarian cancer in even one first degree relative since that is assoc with a 25% chance of having the BRCA 1 mutation (Schwartz et al., 2008)
- any known BRCA mutation in first, second or third degree relatives
- a personal history of irradiation to the breast or chest wall prior to the age of 30, such as for Hodgkin disease, is associated with a very high risk for breast cancer
- Ashkenazi Jewish descent (those tracing their roots to central and eastern Europe) since two BRCA1 and one BRCA2 mutations are observed with higher frequency in Ashkenazi

Jews. Since, taken together, the frequency of these three mutations approximates 1 in 40 in Ashkenazi Jews, we will not enroll women with this ancestral history, unless there is no family history of cancer (in which case the BRCA1/2 mutation frequency is not greater than its normal low prevalence).

It is important to emphasize that the exclusion criteria listed above, based on family history profile, is primarily designed to detect high penetrance genetic heritability of breast cancer - a very low rate phenomenon. Thus, while the planned exclusion criteria are necessarily very conservative to ensure the least possible risk to participants, they are unlikely to exert any appreciable negative impact on our recruitment and enrollment goals. Even the exclusion of Ashkenazi Jews is unlike to detrimentally impact recruitment since fewer than 2.2 % of United States Jews reside in North Carolina.

Study Design, Methods and Procedures

Overview of Protocol: Briefly, medical and psychiatric evaluations will be conducted during an initial Screening period. Eligible women will then undergo initial Baseline Pretreatment assessments, including stress reactivity testing and assessments of baroreceptor sensitivity and endothelial function. Following Time 1 assessments, women will be randomized (in a 1:1 fashion) to either transdermal 17β-estardiol (100 ug/day) or placebo for 12 months. Following randomization, during months 1, 2, 4, 6, 8, 10 and 12, compliance, side effects, vitals, depression symptoms and functional impairment will be assessed. Also, on a quarterly basis, oral micronized progesterone (200 mg/day x 12 days) will be given to ERT treated women. At the end of month 6 (Time 2 testing) and month 12 (Time 3 testing), while estradiol is still on board but prior to the final progesterone administration, all assessments conducted at baseline will be repeated. These events are described below.

D.3. Procedures

I. Screening Medical and Psychiatric Evaluations

Following the informed consent process, subjects will complete a detailed medical and medication history form (enclosed); demographic data and height, weight and waist circumference will be obtained. Questionnaires assessing functional impairment, sleep quality and climacteric symptoms, diet and physical activity level will be administered (enclosed). The CES-D scale will also be administered. CES-D scores ≥ 9 will result in exclusion and referral for further evaluation. Stethoscopic BP will also be taken. Using JNC 7 guidelines, women exhibiting Stage 1 or Stage 2 hypertension (HTN) (140+ mmHg SBP and/or 90+ mmHg DBP) will be excluded and referred. Women will also have a fasting blood draw to determine FSH, plasma lipids (HDL-C, LDL-C, and triglycerides), and fasting glucose. As per the National Cholesterol Education Program (NCEP) and American Diabetes Association guidelines, fasting LDL ≥ 190 mg/dl or fasting glucose ≥ 7.0 mmol/l (126 mg/dl) (exclusionary) will prompt a referral. See A.4.8. for monitoring of change in health status during the course of the study.

Subjects meeting medical and psychiatric screening criteria will have a psychiatric evaluation with clinical psychologist Dr. Forneris, who will administer the Structured Clinical Interview for DSM-IV (SCID-NP with psychotic screen) for Axis I disorders and who will administer the cumulative trauma interview. Any woman meeting criteria for a current Axis I disorder will be excluded and referred for treatment. Women meeting psychiatric inclusion criteria will be scheduled for a gynecological exam with Dr. Steege who will take a complete medical history, including a history of estrogen-dependent neoplasia, a history of thrombophlebitis or thromboembolic disorders, gall bladder disease, liver dysfunction or other disorders for which estradiol or progesterone use is contraindicated. He will also perform a pelvic exam as well as endometrial ultrasound to confirm that they are free of any signs or history of endometrial disorder, abnormal uterine anatomy (e.g. fibroids), abnormal ovarian anatomy (e.g. endometrioma). If the endometrium is > 4 mm, a routine endometrial biopsy will be performed to rule out endometrial cancer and to confirm eligibility for ERT. During the same exam Dr. Steege will also perform a breast exam. All women will be required to have a normal mammogram within one year of study enrollment (and one year thereafter). Screening mammograms will be provided for women who have not had one and one year mammograms provided for all subjects. An abnormal endometrial biopsy or detection of breast lump or abnormal screening mammogram will prompt a referral. Subjects who have a normal follow-up diagnostic mammogram will be allowed to continue in the study.

In order to control as much as possible for extraneous factors that vary with depression and impact cardiovascular risk (e.g., diet and exercise) all subject will have the American Heart Association DASH diet as well as exercise guidelines reviewed with them and each will be encouraged to adhere to them throughout the study (enclosed for review). We will review these with the subject at every study visit. This strategy will also minimize the likelihood that subjects will develop medical conditions (e.g. hypertension, hyperlipidemia) that would result in their withdrawal.

II. BASELINE PRE-TREATMENT ASSESSMENTS (TIME 1)

Subjects will arrive at the laboratory following an overnight fast have fasting blood samples taken for glucose and insulin. A meal will then be provided and subjects will undergo the following laboratory testing procedures.

Laboratory Testing Procedures (see figure)

Baroreflex Sensitivity: While supine. BP will be measured continuously using a noninvasive radial artery tonometric device. The time interval between successive R waves (RR interval) is determined via a standard 3-lead ECG. After 20 min of rest, 5 min of continuous BP and RR interval measurements will be recorded. The spontaneous fluctuations in SBP and RR interval will be analyzed using the Sequence Technique for baroreflex sensitivity (Measurements).

Extended Baseline Rest (40 mins): Following the intravenous (I.V.) setup and instrumentation for cardiovascular monitoring, a 30 min venipuncture recovery period will ensue, followed by a 10 min pre-challenge baseline period for determining cardiovascular, neuroendocrine and inflammatory marker levels. During the pre-challenge baseline, cardiovascular measures will be taken at mins 2, 4, 6, 8 and 10 and averaged. Blood will be sampled at min 10 for norepinephrine (NEpi), ACTH, cortisol, Il-6 and high sensitivity (hs) CRP, and for pre-treatment estradiol and progesterone.

The Trier Social Stress Test (TSST): reliably induces large and consistent HPA-axis, CV, NEpi, and inflammatory responses and reliably differentiates depressed from non-depressed patients. The TSST involves the following: **Pre-Task Instructions**: (5 min) Subjects will be introduced to 3 people (the 'selection committee') after which the experimenter will ask the subject to take over the role of a job applicant who is invited for a personnel interview with the selection committee. **Anticipation Period**: The subject is left alone for 10 mins to prepare. **Speech:** The committee will return to the testing room where the managers ask the subject to deliver her talk for 5 mins. If the subject finishes before 5 mins, the managers respond with prepared questions to ensure that the subject speaks for the entire 5 mins. Immediately following, the committee will ask the subject to subtract the number 13 from 1,022 as fast and as accurately as possible for 5 mins. For each mistake, the subject is instructed to restart at 1,022. **Stress Recovery**: The subject sits alone and quiet for 90 mins.

CV and Neuroendocrine Sampling during Stress Testing:

If primary hypotheses for Stress Reactivity are confirmed then for secondary analyses, we will examine reactivity scores for other measures reflecting sympathetic (cardiac output, pre-ejection period, plasma norepinephrine (Nepi)), and HPA (ACTH and the ACTH cortisol ratio) reactivity, and also basal CRP levels in order to further refine the CVD risk profile. Thus, cardiovascular measures will be taken at min 1,3,5,7, and 9 of the Anticipation Period, min 1,3, and 5 of Speech, and min 1,3 and 5 of serial subtraction and averaged to yield task levels. Cardiovascular measurements will also be taken at 5 minute intervals during the recovery period. NEpi will be sampled at the end of min 2 of Speech delivery and at the end of min 2 of serial subtraction since catecholamines peak within the first mins of stress. Cortisol and ACTH will be sampled during Stress Recovery at min 10, 20 and 30 since HPA-axis responses to stress are reliably found 10 – 30 min after cessation of the TSST, and cortisol at 60 and 90 min to assess recovery. Although the precise time course for IL-6 responses to stress is not definitively determined, IL-6 shows a delayed response following acute stress. We will sample for IL-6 at 10, 30, 60, and 90 min, since other research found that depressed patients differed from controls in IL-6 reactivity to the TSST at 90 min post-stress. There is less evidence for stress-induced increases in CRP, so it will be taken at baseline rest only.

Total blood taken during baseline rest and stress testing = 100 ml (approximately 7 tablespoons)

Under the supervision of Dr. Hinderliter, **Flow Mediated Dilatation of the Brachial Artery** will be assessed by measuring changes in arterial diameter induced by reactive hyperemia (endothelial-dependent) versus nitroglycerin (endothelial-independent). Ultrasound images of the right brachial artery proximal to the antecubital fossa will be acquired. Baseline images will be obtained after 10 min of supine rest. Flow-mediated (endothelial dependent) dilatation will be assessed by determining the change in arterial diameter in response to reactive hyperemia induced by inflating a pneumatic occlusion cuff placed around the forearm to a suprasystolic pressure (approximately 200 mmHg) for 5 min. Images of the artery will be recorded for 2 min after cuff deflation. In order to confirm that any treatment-associated changes in flow mediated dilatation result from enhanced endothelial function, after 10 min of rest, a second baseline image will be acquired. Then, sublingual nitroglycerin spray (0.4 mg) will be administered to determine endothelium-independent vasodilation, and ultrasound images will be acquired for the subsequent 5 min. Flow-mediated and nitroglycerine-mediated dilation will be calculated as percentage changes in diameter from their respective baselines.

III. Randomization and Pharmacological Intervention Procedures

Under double-blind procedures, subjects will be randomized to receive either transdermal 17β-estradiol or placebo (as inert patches) for 12 months, stratified based on histories of recurrent depression (yes, no). The study statistician, Dr. Hamer will be responsible for creating the randomization scheme, and the UNC Hospitals Investigational Drug Services will manage the randomization and dispensing of all study medication in blinded form. The use of specially prepared patches with no outer label will allow both the subject and the experimenter to be blind to the intervention. For those 160 women randomized to receive transdermal estradiol, each will wear a patch (Climara®) designed to deliver 0.10 mg estardiol over a 24 hour period. Transdermal administration of estradiol produces mean serum concentrations of estradiol comparable to those produced by premenopausal women in the early follicular phase. Climara® continuously releases estradiol, which is transported across the intact skin leading to sustained circulating levels of estradiol during each 7-day treatment period. For the 160 women randomized to placebo, each will receive a placebo patch identical in appearance to the active patch (developed by 3M Pharmaceuticals).

IV. 12 Month Intervention Phase

Procedures during the Intervention: A monthly calendar outlining weekly schedules for changing patches, study visits, and the monthly schedule for taking micronized progesterone or placebo progesterone will be provided. Study visits will occur on months 1, 2, 4, 8, and 10 to allow for the careful monitoring of side effects and to ensure compliance. The same assessments will occur on months 6 and 12 as part of the laboratory testing schedule. All visits will occur during the first two weeks of the month, prior to progesterone/placebo administration. During the visits, vitals and side effects will be assessed as will depressive symptoms and functional impairment. SCID interviews will be conducted in those with CES-D scores > 10 to assess minor and major depression. Compliance will be determined (see below) and the next supply of medications will be dispensed. Validated questionnaires will be administered for sleep, climacteric symptoms, diet and physical activity. Quarterly, we will sample for estradiol to confirm compliance at the end of the trial.

Compliance checks (at every study visit): The following strategies will be used to ensure compliance: Subjects will: **1)** be educated on the importance of replacing their patches every 7 days and we will develop an individualized behavioral strategy to increase the likelihood that the subject will do so (e.g., apply new patch every Sunday morning before church); **2)** record date of patch application on their calendars, reviewed at each study visit; **3)** bring their used and unused patches with them to every study visit; **4)** be asked whether they forgot to apply any patches and how many days late they were; **5)** be given extra patches and instructed to check to ensure patches are in place each day and to immediately replace any patch that may have fallen off; and **6)** be instructed that blood will be drawn for E2 in order to assess compliance.

Progesterone Administration: To prevent estradiol-induced endometrial hyperplasia, women randomized to active ERT will take 200 mg micronized progeseterone/day for 12 days every 3 months. Exposure to a progestin every 3 months is sufficient to protect the endometrium (Popp, A.W., et al., *Prevention of postmenopausal bone loss with long-cycle hormone replacement therapy.* Maturitas, 2006. **53**(2): p. 191-200). To help preserve the blind, we will administer progesterone under double-blind procedures. Women randomized to placebo-ERT

will take placebo progesterone each month, while those randomized to active ERT will take active progesterone every third month and placebo progesterone the other months. Thus, all women will take a blinded capsule during the last 12 days of each month. We have chosen to administer progesterone in a discontinuous fashion (which is adequate for safety) as opposed to using Climara Pro® (a transdermal system that delivers both estradiol and a progestin on a continuous basis) because the progesterone in the Climara Pro® system is a synthetic progestin (levenorgestrel) and because progestins antagonize many of estrogen's beneficial CV effects. This will allow for the examination of the effects of estradiol on the mechanisms proposed to mediate the link between depression and CVD without the confound of a continuous progestin.

V. Month 6 and Month 12 Testing (Times 2 and Time 3)

During months 6 and 12, while estradiol is still on board but before the quarterly dose of micronized progesterone is administered, all procedures described above for pre-treatment (Time 1) testing will be repeated, ensuring that the participant is wearing a patch at the time of testing. At the end of the 12 month, Time 3 testing women will be abruptly withdrawn from the intervention since in the majority of women, there is no definitive evidence that tapering the discontinuation of ERT is associated with fewer vasomotor symptoms in the majority of women (Aslan E, et al., *Maturitas*. 2007 Jan 20;56(1):78-83. Epub 2006 Oct 13; Grady D et al., *Am J Med*. 2005 Dec 19;118 Suppl 12B:163-5).

MEASUREMENTS

Cardiovascular Risk Measures

1) Metabolic Risk: We will employ National Cholesterol Education Program (NCEP) standard criteria for determining the Metabolic Syndrome based on blood pressure (BP) levels, abdominal obesity, and fasting triglyceride, HDL and glucose. For **BP**: subjects will be seated for 5 mins with arm supported at heart level. The auscultatory method will be used, employing a hand aneroid sphygmomanometer and stethoscope placed over the brachial artery at the inside of the elbow. A minimum of two BP measurements will be taken at least 5 mins apart and averaged to determine BP. **Waist Circumference** will be measured using the NHANES III Protocol [6]. The subject stands and the examiner palpates the upper hip bone to locate the right iliac crest to determine the mid-axillary line. The measuring tape is placed in a horizontal plane around the abdomen at this point and the measure is made at a normal minimal respiration. Following an overnight fast, blood will be sampled for **fasting triglycerides and HDL** (see assay descriptions below). We will also use **Insulin Sensitivity/Resistance to assess metabolic risk** using Homeostasis Model Assessment (HOMA) beta-cell index derived from a single fasting sample of glucose and insulin levels. This is the most widely used surrogate measure of insulin resistance (IR). Fasting HOMA-IR levels have been shown in large scale studies to predict CV events even after adjustment for other traditional risk factors. We will use the standard calculation for HOMA-IR, which is fasting insulin (uU/ml) times fasting glucose (mmol/l) divided by 22.5. Homo-IR has been validated in normoglycemic subjects against insulin sensitivity measured directly from the euglycemic-hyperinsulinemic clamp technique. Subjects will be considered insulin resistant on the basis of HOMA-IR > 3.2, consistent with other investigations.

Metabolic Risk will be assessed at baseline, month 6 and month 12. Subjects will be classified as having metabolic risk if they either: 1) meet standard NCEP criteria for metabolic syndrome based on any three of the following risk factors: (1) BP ≥130/ ≥85 mmHg, (2) waist circumference > 88 cm, (3) triglycerides ≥ 150 mg/dl, (4) HDL cholesterol < 50 mg/dl, and (5) fasting glucose ≥ 110 mg/dl.; **or 2)** are insulin resistant, a clear metabolic predictor of CVD, i.e., if their HOMA-IR is > 3.2. **2) Sequence Technique for estimating Baroreflex Sensitivity**: 5-min recording of BP and RR interval will be scanned for sequences in which SBP and the subsequent RR interval progressively increase or decrease over 3 consecutive beats. Sequences will be selected if successive pressure pulses differ by at least 1.0 mm Hg and successive RR intervals differ by at least 5.0 msec. Linear regression relating RR interval to SBP will be plotted for each sequence, and the slope of the function will be the estimate of BRS. Slopes derived from all sequences will be pooled so that one measures of baroreceptor sensitivity will be obtained. 3) **Blood Pressure and Vascular Resistance Reactivity to Stress**: The Suntech Exercise BP monitor, Model 4240 (SunTech Medical Instruments, Inc., Raleigh, NC), will provide automated measurement of BP during stress testing using the auscultatory technique. Initially, five standard stethoscopic BPs will be taken simultaneously with the automated pressures in order to insure correct microphone placement and cuff positioning. Impedance cardiography is noninvasive and involves recording changes in thoracic impedance by use of a tetrapolar electrode system, together with a standard ECG. Impedance cardiography has been extensively evaluated by comparison with traditional invasive techniques in human and animal studies. For each min of interest, a 30-s continuous sample of waveforms (obtained concurrently with BP) will be processed to generate an ensemble-averaged cardiac cycle, from which stroke volume will be determined and HR will be determined by the interbeat interval. Cardiac output and total peripheral resistance are calculated using standard formulae. To control for differences in BMI, measures will be adjusted for body surface area, yielding indexed scores (e.g. vascular resistance index (VRI)). **Flow Mediated Dilatation:** Digital gated images of the brachial artery will be acquired for analyses. Measurements will be performed using customized software (Vascular Analysis Tools, Medical Imaging Applications, LLC). Arterial diameter will be quantified by analysis of each image sequence, measuring from the intima-lumen interfaces of the proximal and distal arterial walls. Baseline resting arterial diameters will be quantified as the average of measurements on 10 or more images. Arterial diameters during reactive hyperemia and following nitroglycerin administration will be calculated as the average of measurements on 3 consecutive images at maximum dilatation. Flow mediated and nitroglycerin-mediated dilatation will be expressed as percent changes in diameter relative to resting scans.

Affective Well-Being Measures

1. Depression Episodes will be operationalized in a manner consistent with other studies in perimenopausal depression [7] and based on the following assessments: **The Center for Epidemiologic Studies Depression scale (CES-D)** is a 20-item self-report form assessing depressed mood. CES-D scores ≥ 16 have been consistently validated as defining high depressive symptoms suggestive of clinically significant DEP (positive predictive value=75%), and this cut-point has been used to define depressive episodes in multiple perimenopausal samples. Moreover, elevated CES-D scores have been shown to be sensitive to the anti-depressant actions of ERT in PM women. **Thus, CES-D scores ≥ 16 will be used to operationally define 'depression episodes'.** However, if CES-D scores are > 10 at any study visit then the **Structured Clinical Interview for DSM-IV** (non-patient edition) will be

administered to determine the presence of either minor or major depression. **Functional Impairment:** We will use the widely used Medical Outcomes Study 36-item short form (SF-36), which yields scores of general health perceptions, physical functioning, role limitations due to physical functioning, social functioning, bodily pain, general mental health, role limitations due to emotional problems, and vitality. The rationale for the selection of this instrument is based on: 1) the demonstration that it is sensitive to change over time; 2) its incorporation of subscales to specifically assess functional impairment related to emotional and physical symptoms as well as overall social functioning; 3) its able to discriminate subgroups of CVD patients known to have poorer survival and worse disease; and 4) normative data for this scale are available for the general US population as well as 13 medical conditions**. Symptom Severity and Affective Instability** will be assessed with the CES-D scale (at baseline and months 2,4,6,8,10 and 12), the former with a repeated measures approach and the latter by the variance in the CES-D score over 12 months.

Lifetime Trauma Exposure will be assessed by a validated interview. A cumulative (total) trauma score will be generated by assigning one-point for each of the following 15 traumas: **1)** history of sexual abuse; **2)** history of physical abuse; **3)** moderate or greater childhood physical neglect; **4)** moderate or greater childhood emotional neglect; **5)** presence of 7 other types of trauma (one point each) before the age of 18; **6)** presence of 4 lifetime traumas (1 point each). The measure of lifetime ***sexual and physical abuse history*** is based on previous studies by Investigator, Dr. Leserman, and adapted from other research. Cumulative trauma exposure has been shown to independently predict the development of adverse medical conditions, to be related to a variety of risk factors for the leading causes of death; and to predict CV reactivity to acute laboratory stressors.

Benefits to Subjects and/or Society

Subjects randomized to active ERT may experience a direct benefit, as treatment may reduce depressive episodes, improve quality of life, and reduce cardiovascular risk. The benefits to all subjects also include study-related medical evaluations, including a gynecological exam and mammograms. Subjects may also benefit by knowing that they are contributing to research aimed at enhancing our understanding of depression and cardiovascular risk in perimenopausal women. Even women who are excluded from participating based on medical or psychiatric criteria are intended to benefit to some degree since they will receive valuable feedback on their current medical and psychiatric status and referral for treatment when indicated. Society may benefit if this study yields knowledge on predictors of ERT benefit in perimenopausal women.

Risks and Measures to Minimize Risks

The most likely risks involve side effects associated with the use of reproductive hormones (estradiol or progesterone).

The most frequent side effects associated with estradiol use include edema, breast tenderness, and changes in appetite and weight. Less frequent side effects include jaundice, nausea, abdominal cramps, increased blood pressure, headache and worsening of migraines or asthma, depression, nervousness, acne, cystitis-like syndrome, enlargement of uterine fibroids, intolerance to contact lenses and dizziness. Skin rash or irritation may also occur at site where the patch (containing estrogen) is placed. Progesterone - The most common side effects

associated with progesterone include breast tenderness, dizziness, abdominal bloating, vaginal discharge, chest pain, and diarrhea. Less common side effects include headache, dizziness, breast pain, musculoskeletal pain, and viral infection

More Rare yet Serious Risks:

The risk of **venous thromboembolism** (VTE) is clearly increased with HRT, with the risk influenced by age and preparation/type therapy. In the Women's Health Initiative Trial, the risk of VTE was greater with combined estrogen + progestin (HRT) (HR 2.06) than with conguated equine estrogen alone (HR 1.32), with both risk rates reduced in the 50-59 year olds (estimated excess events 9-10/1,000 for HRT and 3-4/1000 for ERT) (Rossouw, J.E., et al., *Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial.* Jama, 2002. **288**(3): p. 321-33. Lyytinen, H., E. Pukkala, and O. Ylikorkala, *Breast cancer risk in postmenopausal women using estradiol-progestogen therapy.* Obstet Gynecol, 2009. **113**(1): p. 65-73).

Notably, the risk for VTE was significantly reduced by the use of transdermal estradiol in case control studies (HR 0.9 compared with HR 4.2 in those taking oral estrogen) (Scarabin, P.Y., et al., *Effects of oral and transdermal estrogen/progesterone regimens on blood coagulation and fibrinolysis in postmenopausal women. A randomized controlled trial.* Arterioscler Thromb Vasc Biol, 1997. **17**(11): p. 3071-8.

Both HRT and ERT increase the risk of **stroke** Hendrix, S.L., et al., *Effects of conjugated equine estrogen on stroke in the Women's Health Initiative.* Circulation, 2006. **113**(20): p. 2425-34). Not surprisingly, the increased risk is age-dependent (excess attributable risk 1 case per 5000 women years in the 50-59 year olds vs 4.5 in the whole Women's Health Initiative Study group), largely reflecting the increased risk of stroke with aging.

Breast Cancer – Both HRT and ERT increase the risk of breast cancer (although the impact of duration and timing of therapy are still uncertain). Focusing on ERT (which decreases the risk compared with HRT by almost a factor of four) (Beral, V., *Breast cancer and hormone-replacement therapy in the Million Women Study.* Lancet, 2003. **362**(9382): p. 419-27),an early meta-analysis concluded that no increased risk of breast cancer occurred if estrogen alone was administered for less than five years (*Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer.* Lancet, 1997. **350**(9084): p. 1047-59).

A subsequent meta-analysis reported a relative risk increase of 3.1%/year [8] Taking the worst case scenario, the risk in a 50-54 year old woman who takes estradiol alone for five years would increase from 13/1000 women (with no ERT) to 14.94/1000 women (Endocrine Society, in press). Thus, the risk of exposure to one year of ERT would be insubstantial. **Ovarian cancer** – Long term use of ERT alone increases the risk of ovarian cancer, with an attributable risk of 0.7 per 1000 women per 5 years of use (Lacey, J.V., Jr., et al., *Menopausal hormone replacement therapy and risk of ovarian cancer.* Jama, 2002. **288**(3): p. 334-41.

326. Morch, L.S., et al., *Hormone therapy and ovarian cancer.* Jama, 2009. **302**(3): p. 298-305).

Endometrial Cancer - Exposure to unopposed estrogen increases the risk of endometrial cancer two fold (Grady, D., et al., *Hormone replacement therapy and endometrial cancer risk: a meta-analysis.* Obstet Gynecol, 1995. **85**(2): p. 304-13). Use of a progestin prevents both endometrial hyperplasia and the increased risk of endometrial cancer (Lethaby, A., et al., *Hormone replacement therapy in postmenopausal women: endometrial hyperplasia and irregular bleeding.* Cochrane Database Syst Rev, 2000(2): p. CD000402).

Although sequential use of progestin does not protect the endometrium to the same degree as continuous combined, long cycle sequential progestin (q 3 month progestin) (as is proposed for use in our study) has not been associated with increased risk of endometrial cancer (Furness, S., et al., *Hormone therapy in postmenopausal women and risk of endometrial hyperplasia.* Cochrane Database Syst Rev, 2009(2): p. CD000402; Odmark, I.S., et al., *Endometrial safety and bleeding pattern during a five-year treatment with long-cycle hormone therapy.* Menopause, 2005. **12**(6): p. 699-707).

Risks Associated with Brachial Artery Imaging: There are no significant risks associated with the ultrasound assessment of brachial artery imaging, though the inflation of the pneumatic cuff to 200 mmHg for 5 mins may be painful. Side effects associated with the sublingual administration of nitroglycerine include headache, dizziness, flushing or restlessness.

Psychological Risks: The structured interviews to assess current and lifetime psychiatric illness and prior trauma may be associated with some psychological distress. Some items in the questionnaires may provoke some negative emotion in some individuals. There is also the risk a woman will experience the onset of a depression episode or have a worsening of depressive symptoms over the course of the 12-month intervention period that would go undetected or untreated.

Standard risks associated with venipuncture include syncope, discomfort, nausea, bruising. Although measures are taken to protect the privacy of every participant, there is a remote risk of breach of confidentiality/loss of privacy.

Protection Against Risk:

- Risks associated with Estrogen and progesterone are minimized in our protocol by i) including only young, medically healthy perimenopausal women undergoing a natural menopause transition; ii) a thorough gynecological exam and medical history conducted by study gynecologist Dr. Steege to ensure that all women are eligible candidates for ERT; 3) the use of intermittent micronized progesterone (200 mg/day for 12 days every 3 months) to eliminate any increased risk of endometrial disorder associated with ERT and to avoid exposure to other progestins; 4) the requirement for a normal mammogram and breast exam at study entry; 5) the exclusion of women with a family history indicative of increased breast or ovarian cancer risk (see exclusion criteria in A.4.4. above); 5) exposure to estradiol for only 12 months since short-term exposure (< 3-5 yrs) has not been shown to be associated with an increased risk of breast cancer; 6) the frequent assessment of vitals, side effects and depression symptoms (months 1,2,4,6,8,10 and 12); 7) the provision of instruction and instructional sheet (enclosed for review) for performing monthly self-breast exams; 8) provision of a patient information sheet with the instruction to call the study doctors immediately if any of the following symptoms are

experienced: severe or prolonged vaginal bleeding, severe headaches, changes in vision or speech, mental depression, nervousness, or suicidal thoughts, dizziness or faintness, pains in calves or chest, sudden shortness of breath, or detection of a breast lump; (enclosed for review); 8) the development of a safety monitoring plan (see below); and 9) monitoring for adverse events by a Data and Safety Monitoring Committee.

While risks associated with the endothelial flow mediated dilatation assessment may include discomfort or pain associated with inflation of the pneumatic cuff, the pain lasts only a few minutes and there is no risk of tissue damage during 5 minutes of forearm ischemia. Moreover, numerous studies in our laboratory using the submaximal effort tourniquet procedure to assess voluntary pain tolerance following inflation of a pneumatic arm cuff to 200 mmHg have reported that the average voluntary tolerance for women is approximately 650 seconds (> 10 mins), twice as long as the time required in the present study. Risks of dizziness associated with nitroglycerine will naturally be minimized since subjects will be in the supine position and instructed to stand up slowly. While headache may develop, our prior experience with this procedure is that headaches were mild, short-lived, and did not interfere with function.

Psychological risks associated with the interviews and stress testing are minimized via the use of clinically trained interviewers and highly competent and professional staff. Frequent contact with subjects and frequent assessment of depressive symptoms (months 1, 2, 4, 8, 10, and 12) will minimize the risk that a worsening of depressive symptoms would go undetected. Risks associated with phlebotomy will be minimized by using sterile materials, by applying pressure to the skin upon removal of the needle, and by placing each subject in the reclined position to minimize risk of syncope. Risks of loss of confidentiality will be minimized by the use of subject codes linked to all data and by maintaining the file linking the subject code to the name in a separate secure location. All databases will be stored on a secure, HIPAA compliant network. The server is backed up nightly. The database is also protected by logins and passwords required to open the system. Different levels of privileges are granted to different users. This study will use a web-based, secure, password protected access to the database. Passwords will be changed every 90 days.

Data Monitoring and Analysis

The safety of the study will be monitored by a **Data and Safety Monitoring Board (DSMB).** This will ensure the independent oversight of issues related to safety and adverse events. This DSMB will consist of 4 members: a Cardiologist, a Psychiatrist, a Biostatistician, and a Gynecologist.

I. Grading of Adverse Events (AEs)

Adverse events will be classified according to severity as either mild, moderate, or severe.

Mild AE: is defined as having no effect on activities of daily living such as transient light headedness or sweating with venipuncture; breast tenderness, mild skin irritation; or something of equal significance that requires no medical intervention and is of marginal clinical relevance.

Moderate AE: would be associated with temporary (minutes to a few days) disruption in activities of daily living such as temporary loss of consciousness with venipuncture; a worsening of migraines or headache that require bed rest, an increase in depression symptoms consistent with mild DEP mood (CES-D scores = 10 - 15), excess sleepiness or fatigue that resolves with additional rest, or something of equal significance.

Severe AE: would cause serious disruption in daily activities of living and may or may not require hospitalization. Examples of such events would include a thromboembolic event, breast carcinoma, an increase in blood pressure consistent with Stage 2 hypertension, any cardiovascular event, any event that is permanently disabling, any event requiring hospitalization or is life threatening, severe mood impairment or failed suicide attempt or something of equal significance. Death would also be considered a severe AE.

II. Monitoring and Reporting of Adverse Events (AEs)

The study coordinator will monitor side effects and AEs at all study visits (on a monthly basis initially and then quarterly) and report all side effects as well as any significant mood deterioration (see below) directly to Dr. Girdler following study visits. Side effects consistent with moderate or severe AEs or that are troublesome to the subject, or evidence of mood deterioration reported by telephone, will also be immediately reported to Dr. Girdler. Dr. Girdler will ensure that Dr. Rubinow or Dr. Steege are consulted within 24-hours. Staff will be specifically trained to monitor the CES-D rating scales for scores > 10 or for endorsement of suicidal ideation and report any such occurrence immediately. In the case of mood deterioration, Dr. Rubinow will determine the next course of action, including a reassessment of mood in two weeks or more urgently if deemed to be clinically required. Dr. Steege will be the responsible physician to determine the next course of action involving medical side effects (e.g., continue in protocol, exclude and refer for treatment).

The PIs, Dr. Rubinow and Girdler will review all protocol data at monthly meetings, including enrollment and retention statistics and aggregate reports of side effects/AEs. As the contact PI, Dr. Girdler will be the one responsible one to report any individual occurrence of an AE to the Chair of the DSMB according to the guidelines established at the initial DSMB meeting (e.g. the DSMB may require the reporting of any serious adverse event on an individual basis and in an on-going fashion). She will also report all moderate or severe AEs to the IRB and CTRC within 1 week. Since we are employing a marketed pharmaceutical product (i.e., a non-IND study), unexpected Serious AEs will be also be reported to the FDA Medwatch Program. The NIMH program officer will be notified of any study modifications or suspension imposed by the DSMB or local IRB in response to an AE.

Dr. Girdler will also report all AEs graded as moderate or severe to the study gynecologist, who will maintain a secured, password protected file linking subject ID to subject name and will also

receive the randomization code by subject ID from Investigational Drug Services. In the event that it is medically necessary to become unblinded to treatment assignment, he will be able to do so very quickly.

We also have a plan for the monitoring of medical conditions that may emerge during the course of the 12 month study requiring treatment or withdrawal. The development of hypertension or hyperlipidemia represent the most likely medical conditions that may emerge over a 12 month interval in perimenopausal women. Our protocol allows for the frequent monitoring of BP levels (months 1, 2, 4, 6, 8, 10, and 12) throughout the protocol. Consistent with JNC 7 guidelines, the detection of BP levels during the study in the Stage 1 HTN range (140 – 159 mmHg SBP or 90 – 99 mmHg DBP) will prompt monthly reassessments of BP to confirm BP status, and a review of the American Heart Association (AHA) dietary and physical activity guidelines with the subject. If the average BP over any 3 consecutive months remains in the Stage 1 HTN range, the subject will be withdrawn and referred for treatment. Detection of BP levels consistent with Stage 2 HTN at any point will prompt immediate withdrawal and referral. Fasting LDL cholesterol will be measured at months 6 and 12 (Time 1 and 2 testing). However, in women with LDL-cholesterol > 190 mg/dl at month 6, a level for which the NCEP recommends therapeutic lifestyle changes and LDL-lowering drugs, we will re-assess fasting lipid profiles at months 8 and 10. Should LDL again exceed 190 mg/dl on either determination, the subject will be withdrawn from the study and referred for treatment. If a subject develops coronary heart disease, venous thromboembolism, breast or endometrial or ovarian cancer, or biliary tract disease, she will be withdrawn from the study. Our plan to enroll only medically healthy perimenopausal women and to educate, encourage, and assess compliance with the AHA diet and exercise guidelines will minimize the development of medical conditions requiring withdrawal.

General Statistical Considerations: This is a two between subjects and one within-subjects (repeated measure) design. The two between-subjects factors are treatment (ERT or placebo, randomized) and history of DEP (positive or negative, non-randomized), and the within-subjects factor is time (psychiatric outcomes at months 0 (baseline), 1, 2, 4, 6, 8, 10, and 12, and CV measures at months 0, 6, 12). Analyses which use outcomes measured repeatedly will use this repeated measures design for analysis, with mixed models for continuous outcomes and logistic models for categorical outcomes. The mixed models we will fit will be random coefficient models, which treat the time variable as numerical. We do this for several reasons: (1) it allows for a finer assessment of precisely when women come in for a "visit," as visits rarely take place on exactly the day planned, and (2) the covariance structure is necessarily simpler, leading to more efficient mixed models. For comparisons assessing change over time, we will assess this change as a difference between treatments or groups, as stated with respect to specific hypotheses, on differences between baseline and month 12 least squares means (LSMs). The month 6 CV data will be part of the mixed model predicting the 12 month least squares means. That is, the mixed model is an Intent- to-Treat analysis which uses all available data to make a prediction about what outcomes we would see if all subjects had remained in through the 12 month endpoint. Several outcomes are measured or derived only once over the follow-up period (e.g., presence of at least one depressive episode during follow-up, or affective instability during the follow-up period). For those outcomes, the model will not include a repeated factor, and will be a two-factor ANOVA or logistic regression. For each analysis, where appropriate, we will perform sensitivity analyses by adding potential mediators, moderators, or confounders (see

D.4.3.) in an appropriate way to the model to assess the effect of these confounders, and to better define mechanisms, and subjects for whom, for example, treatment is effective. Most hypotheses will be examined as main effects, interactions, or contrasts. Prior to statistical analyses, we will use descriptive statistics and graphics to screen the data for errors, outliers, and potential influential observations, and to check distributional assumptions. We will not blindly switch to non-parametrics if distributions within groups do not appear to be normal, but will judge whether the type and degree of non-normality are such which require rethinking our modeling assumptions. We have a relatively large sample size and our linear models should be robust to relatively mild violations of normality. Even so, in the event we decide that the linear or linear mixed models we've selected are not appropriate, we will examine the possibility of testing hypotheses using permutation tests, sampling from the permutation distribution, in preference to non-parametrics.

We will define *affective instability* as the variance of each subject's CED-D total score over her follow-up period. We define the *CV risk* variable S*tress Reactivity,* as a composite formed in the following way: for IL-6, Mean Arterial Pressure, Vascular Resistance, and Cortisol, we will form the difference between resting and stress measures, yielding *reactivity* for each measure. We will calculate mean and SD for each measure at baseline (pre-treatment) and use those to standardize each measure at baseline (pre-treatment) and during the intervention to a common scale. We will then add the z-scores for the 4 measures together to form our stress reactivity composite at each point measured.

Multiple Testing: Several of the hypotheses use multiple outcome measures. We have in each identified one outcome as the primary outcome variable, and testing on this variable will be done without correction for multiple comparison. We will only examine other variables if the null hypothesis is rejected for the primary outcome variable. This enables us to follow significant effects with other analyses to explain or refine the response, helping us interpret the results.

SPECIFIC AIMS AND HYPOTHESES:

Aim 1: To confirm that PM women with histories of recurrent DEP will, at baseline, suffer from greater CV risk relative to never DEP PM women, and to examine the role of cumulative trauma exposure in the relationship of history of recurrent DEP and CV risk.

H1: When compared with never DEP women, women with histories of recurrent DEP will exhibit a greater Stress Reactivity composite score . If confirmed, the following parallel secondary hypotheses will be examined in hierarchical order: Women with histories of recurrent DEP will exhibit: 1) an increased rate of metabolic risk; and 2) reduced endothelial flow-mediated dilatation and baroreceptor sensitivity.

H2: Secondary Hypothesis: Cumulative trauma exposure will moderate the relationship between histories of recurrent DEP and CV risk such that greater trauma exposure will explain part of the relationship between DEP history and elevated CV risk (tested with sensitivity analyses).

Aim 2: To examine progression in CV risk and affective well-being over 12 months in untreated PM women as a function of histories of DEP.

In the 160 women randomized to placebo we will test the following hypotheses by comparing women with a history of DEP with never depressed women for changes over 12 months:

H1: Relative to baseline reactivity, untreated women with a history of DEP will show a greater increase in the Stress Reactivity composite score over 12 months compared with never DEP women. If confirmed, the following parallel hypotheses will be examined in hierarchical order: In women with a history of DEP: 1) there will be a greater increase in rate of metabolic risk; and 2) a greater reduction in endothelial flow-mediated dilatation and baroreceptor sensitivity.

H2: A greater proportion of untreated women with a history of DEP will exhibit episodes of DEP over 12 months compared with untreated women without a history of DEP. If confirmed, the following parallel hypotheses will be examined in hierarchical order: Untreated women with a history of DEP will have 1) higher CES-D DEP scores; 2) increased functional impairment; and 3) increased affective instability (greater variance in CES-D over 12 months).

Aim 3: To examine predictors and mediators of ERT- related treatment efficacy for affective well-being and CV risk.

In the 320 women randomized to treatment we will test the following hypotheses:

H1A: Women on active treatment will show a <u>decreased</u> Stress Reactivity composite score over 12 months relative to baseline levels and relative to the change (increased stress reactivity) in women on placebo. If confirmed, the following parallel hypotheses will be examined in hierarchical order: Women on active treatment will show 1) a lower rate of metabolic risk; and 2) increased endothelial flow-mediated dilatation and baroreceptor sensitivity. These hypotheses will be tested as main effects of treatment.

HIB: A history of DEP will predict a greater effect of ERT on Stress Reactivity: i.e., a greater decrease in Stress Reactivity (relative to baseline) between active and placebo treated women will be seen in women with a history of DEP (due to their higher baseline reactivity) compared with never DEP women. If confirmed, the following parallel hypotheses will be examined in hierarchical order: a history of DEP will predict a greater effect of ERT for 1) reductions in rate of metabolic risk, and 2) increased endothelial flow-mediated dilatation and baroreceptor sensitivity. These hypotheses will be tested as interactive effects (history by treatment).

H2: Although no subjects will be depressed at baseline, active treatment will be associated with a lower proportion of women experiencing episodes of DEP over 12 months relative to placebo treated women. If confirmed, the following parallel hypotheses will be examined in hierarchical order: Active treatment will be associated with: 1) lower CES-D DEP scores; 2) decreased functional impairment; and 3) decreased affective instability.

H3: <u>Secondary Hypothesis:</u> Treatment-related reductions in Stress Reactivity will partially mediate the beneficial effect of E2 on DEP episodes. If confirmed, the following parallel hypotheses will be examined in hierarchical order: Treatment reductions in Stress Reactivity will partially mediate reductions in 1) CES-D DEP rating scores; 2) Functional Impairment; and 3) Affective Instability (tested with sensitivity analyses).

DATA ANALYTIC PLAN FOR EACH HYPOTHESIS

Aim 1 Hypothesis 1: Since subjects have not yet been assigned to treatment, we will treat analyses of baseline data as two-group analyses rather than a main effect in a two-by-two factorial. The Stress Reactivity composite score as the response. If the null hypothesis is rejected, we will examine the same contrast for metabolic risk (as a parallel logistic regression model since it is categorical), flow mediated dilatation, and baroreceptor sensitivity.

Aim 1 Hypothesis 2: This secondary analysis will add the effect of trauma to the model, and assess whether trauma moderates the effect of DEP history on CV risk variables (above).

Aim 2 Hypothesis 1: In the mixed random coefficients model we will use for this repeated measures analysis, we will test this hypothesis using a contrast that examines the difference between the women with and without histories of DEP on the difference between the month 0 and month 12 least squares means. This contrast will be done in the context of the full model but involves means from 2 of the 4 between-subjects cells. If this null hypothesis is rejected, we will test the same hypotheses for flow mediated dilatation and baoreceptor sensitivity (mixed model), and examine metabolic risk (categorical) via parallel logistic regression model.

Aim 2 Hypothesis 2: We will fit a two-factor logistic regression predicting the presence or absence of DEP episodes during the follow-up period. If we reject the null hypothesis, we will follow this with a two-factor mixed random coefficient model using CED-D, functional impairment and (using without the repeated measures portion since affective instability is defined over the entire follow-up period) affective instability as defined earlier. Further, as sensitivity analyses, we will examine the same effect on number of episodes / month during the follow-up period, using Poisson regression, and on the CES-D scores themselves using a mixed model.

Aim 3 Hypothesis 1 (A and B): With stress reactivity as the response, we will test these hypotheses in the context of the two between-factor one-repeated factor mixed random coefficient model we listed earlier. Hypothesis H1A is the main effect for treatment, predicting that the difference between the change in LSMs from month 0 to month 12 will be in different directions in the treated and untreated women. If this null hypothsis is rejected, we will examine parallel effects on metabolic risk (via a logistic regression), baroreceptor sensitivity, and flow mediated dilatation. Hypothesis H1B involves partitioning the interaction effect in these models.

Aim 3 Hypothesis 2: This will be examined in a logistic regression model with two between-subjects factors (DEP history and treatment) and the emergence of at least one DEP episode as the response. This is the treatment main effect in the model. If we reject the null hypothesis, we will examine the same hypothesis using CED-D total (mixed model), functional impairment (mixed model), number of episodes (Poisson regression), and affective instability (mixed model).

Aim 3 Hypothesis 3 (secondary): This hypothesis will use a logistic regression model with two between-subjects factors as above, the emergence of at least one DEP episode as the response, and stress reactivity in the model as a mediator.

POWER AND SAMPLE SIZE CALCULATIONS FOR EACH SPECIFIC AIM:

It is not feasible to conduct power analyses for every hypothesis. While this is a mechanistic study designed, in part, to examine mediators of E2's beneficial effects on CV risk and DEP, we must first be able to detect a beneficial effect of E2. Thus, we conduct power for Aim 3, H2. We also felt it important, however, to confirm we are powered to detect meaningful differences in the Stress Reactivity composite score since it is our primary CV outcome variable and also a predicted mediator.

Aim 3 Hypothesis 2: Based on the work of Freeman et al (Menopause, 2009. **16**(4): p. 728-34) for rates in untreated women, and our prediction that ERT will cut the rate of depression by 50% (a clinically relevant difference), we predict the following percentages of women will experience at least one DEP episode during the follow-up in our groups:

		Treatment Assigned	
		ERT	**Placebo**
Depression History	**Positive**	**22.5**	**45**
	Negative	**15**	**30**
	(Combined)	**(18.75%)**	**(37.5%)**

The hypothesis corresponds to using logistic regression to test the depressive episode rate between all those treated with ERT (18.75%) versus those treated with placebo (37.5%). Using a logistic regression with this effect size, a two-tailed significance level of 0.05, and 240 total subjects, power exceeds 96%. Further, power should be higher for the parallel poisson regression on number of episodes, and for the analysis of the CES-D scores themselves.

Aim 1, H1: Our primary hypothesis for Aim 1 involves Stress Reactivity and since we cannot assert that any stress-responsive factor is more important than another in this context, we will combine different measures of stress-responsive CV risk factors and combine them into an index called Stress Reactivity, described above. Since this measure has not been used before, we do not have information from prior research on means and standard deviations which could help with power and sample size calculations. Instead, we performed power calculations for each of the stress reactivity variables which comprise the composite score. We used differences between DEP and non-DEP groups from Pace et al. (Am J Psychiatry, 2006. **163**(9): p. 1630-3) for IL-6; from Young et al. (Neuropsychopharmacology, 2000. **23**(4): p. 411-8) for cortisol; from Brownley et al. (Am J Obstet Gynecol, 2004. **190**(4): p. 1052-8) for clinically relevant differences in MAP; and from Girdler (Health Psychol, 2007. **26**(2): p. 201-13; Int J Psychophysiol, 1994. **17**(3): p. 233-4)8).for group differences in VRI – using the reported group differences to define the magnitude of differences which are important, and to obtain estimates of SDs. We obtained effect sizes of 0.63, 0.86, 10, and 1.5 for differences in delta MAP, VRI, IL-6, and Cortisol, respectively. Our sample size of *160* per DEP group yields **99% power to reject a null hypothesis** of no effect for an effect size of 0.63, (the smallest above), and 80% power to detect an effect size as small as 0.40. Due to the complexity of the random coefficient

model to be used for this variable, we approximated our calculations with a simpler ANOVA power on change scores. Our actual power should be higher because the model should be more precise.

Inducements for Participation

Total possible monetary inducement for participating = $1,275, prorated as follows:

$100 for the medical and psychiatric screening (these screenings also provide benefit in the form of free medical evaluations)

$150 for each laboratory test session ($450 total)

$100 for each flow mediated dilatation procedure ($300 total)

$ 25 for each interim study visit ($125 total)

$300 compliance bonus for completing the study in full compliance