|  |
| --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT393 |
| **Submission Date** | May 21, 2020 |
| **Project Title** | Genome-wide meta-analysis for post-partum depression |
| **Tentative Lead Investigator** *(first author)* | Jerry Guintivano (University of North Carolina at Chapel Hill) |
| **Tentative Senior Author** *(last author)* | Jerry Guintivano |
| **eMERGE Site Sponsor & Contact** | Shefali Setia Verma, Marylyn D. Ritchie, University of Pennsylvania |
| **All Other Authors**  | TBD |
| **Sites Participating** | TBD |
| **Background / Significance** | Post-partum depression (PPD) refers to depression experienced by women during pregnancy and up to 12 months after giving birth. The perinatal period captures disorders during pregnancy and postpartum. It is estimated that approximately 10-20% of women that get pregnant or give birth experience depression related to pregnancy and childbirth which makes it one of the most prevalent psychopathologies. As of June 2019, there has only been one genome-wide association study (GWAS) published on postpartum depression (PPD), and it was underpowered. Because this condition *directly* affects mothers within one year of delivery, but *indirectly* affects other parents and all children in a family, it is critical that we spend resources to uncover the genetic architecture of PPD so that new treatments and prevention strategies can be developed. Electronic Health Records (EHRs) linked to a biobank offer a unique opportunity to identify a large group of women who have delivered babies, some of whom will have PPD recorded in their EHR and some who will serve as unaffected controls. An EHR provides an extensive report of the overall health and disease spectrum of every individual within a health care system. EHRs contain a vast amount of information in both structured (diagnosis codes, clinical lab measures, medications, questionnaires) and unstructured (patient notes) data. With broad adoption of EHRs over the past ten years, the research community has begun to leverage this rich, longitudinal data for research purposes including population health and human genetics research. Over 250 GWAS have been published on psychiatric disorders such as major depression, bipolar depression, and schizophrenia to name a few; this is part of our surprise that there has been only one underpowered GWAS on Postpartum Depression. GWAS in other psychiatric disorders have identified hundreds of susceptibility loci which enables a wealth of downstream research for developing new treatment and prevention approaches. As with most psychiatric disorders, PPD is thought to be due to both genetics and non-genetic factors. Thus, in this proposal, our goal is to explore both genetic factors that explain the underlying architecture of PPD.  |
| **Outline of Project** | We will define post-partum case and control sample based on ICD codes and medication data and then run GWAS in eMERGE imputed data to then share with UNC group for meta-analysis. |
| **Desired Data - Common Variables\*** *(Available from the CC)* | X Demographics  X ICD9/10 codesX CPT codes* Phecodes

X BMI | * Common Variable Labs

X Common Variable Meds* Other: Case/Control status on Phase I and Phase II phenotypes
 |
| **Other Desired Data *(Available from participating sites)*** | *Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)*  |
| **Desired Genetic Data** |  X eMERGE I-III Merged set (HRC imputed, GWAS)* eMERGE PGx/PGRNseq data set
* eMERGEseq data set (Phase III)
* eMERGE Whole Genome sequencing data set
* eMERGE Exome chip data set
* eMERGE Whole Exome sequencing data set
* Other (not listed above):
 |
| **Does project pertain to an existing eMERGE Phenotype?** | * Yes, if so please list

 X No |
| **Planned Statistical Analyses** | Genome-wide association study. Logistic Regression adjusting for age, PCs |
| **Ethical Considerations** | N/A |
| **Available Funding or Resources** |  |
| **Target Journal** | TBD |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Analysis to complete by April 2020 to then share with UNC. |

**\*Common Variables available across all datasets:**

* Demographics: sex, year of birth, decade of birth, race, ethnicity
* Codes: (repeated values & age at event): ICD, CPT, Phecodes
* BMI: (repeated value & age at event) height, weight, BMI
* Labs: (lab name, repeated lab value & age at event) Serum total cholesterol, LDL, HDL, Triglycerides, Glucose fasting/non-fasting/unknown, & White Blood Cell count
* Medications: (medication name, repeated, & age at event) Cerivastatin sodium, Rosuvastatin, Simvastatin, Fluvastatin, Pravastatin, Lovastatin, Atorvastatin, & Pitavastatin
* Other: Case/Control status on Phase I and Phase II phenotype: only on GWAS dataset participants