An Intersectoral Approach to Study Built Environment Factors Affecting Postpartum Depression and Children's Health

Center for Transportation, Environment, and Community Health
Final Report

by
Yiye Zhang, Alison Hermann, Rochelle Joly, Elizabeth Feuille, Jyotishman Pathak

March 31st, 2021

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2.Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle An Intersectoral Approach to Study Built Environment Factors Affecting Postpartum Depression and Children's Health		5. Report Date March 31st, 2021 6. Performing Organization Code
7. Author(s) Yiye Zhang, Alison Hermann, Rochelle Joly, Elizabeth Feuille, Jyotishman Pathak		8. Performing Organization Report No.
9. Performing Organization Name and Address Weill Cornell Medicine, Cornell University 425 E. 61st St, Suite 301, New York, NY		10. Work Unit No. 11. Contract or Grant No. 69A3551747119
12. Sponsoring Agency Name and Address U.S. Department of Transportation 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered Final Report 10/01/2019– 03/31/2021 14. Sponsoring Agency Code US-DOT
15. Supplementary Notes		

16. Abstract

Project Abstract

This study examined the association of the quality of built environment in the neighborhood and maternal mental health. In Aim 1 of its 3 aims, we developed predictive models to predict and subtype postpartum depression (PPD) using clinical and built environment information. The predictive model was constructed using a combination of machine learning models, and achieved an AUC of over 0.9 in prediction. The subtyping work identified 3 subgroups of women based on their risk of PPD. Women who experienced higher rates of PPD were more likely to reside in neighborhoods with homogeneous land use, lower walkability, lower air pollutant concentration, and lower accessibility to retail stores after adjusting for age, neighborhood average education level, marital status, and income inequality. In Aim 2, interviews were conducted with clinicians, including obstetrics-gynecologist, reproductive psychiatrist, and pediatricians, on their views of how to leverage information on neighborhood built environment in routine clinical care. In Aim 3, we constructed a list of variables representing social determinants of health including built environment across the states of New York, New Jersey, Connecticut, and Pennsylvania. Using these, a univariate and multivariate analysis was conducted with mothers' PPD and children's allergy status as the outcomes and the environmental variables as predictors. Outputs of the overall study includes journal publications (including being named as Editor's Choice at the Journal of Affective Disorders), invited presentation at Epic Corporation, and findings that lead to ongoing preparation for a clinical trial at Weill Cornell Medicine.

17. Key Words	18. Distribution Statement
Electronic Health Records, Built Environment, Maternal Mental Health	Public Access

19. Security Classif (of this report)	20. Security Classif. (of this page)	21. No of Pages	22. Price
Unclassified	Unclassified		

Form DOT F 1700.7 (8-69)

Aim 1.1 Development and Validation of a Machine Learning Algorithm for Predicting the Risk of Postpartum Depression among Pregnant Women

Abstract

Objective

There is a scarcity in tools to predict postpartum depression (PPD). We propose a machine learning framework for PPD risk prediction using data extracted from electronic health records (EHRs).

Methods

Two EHR datasets containing data on 15,197 women from 2015 to 2018 at a single site, and 53,972 women from 2004 to 2017 at multiple sites were used as development and validation sets, respectively, to construct the PPD risk prediction model. The primary outcome was a diagnosis of PPD within 1 year following childbirth. A framework of data extraction, processing, and machine learning was implemented to select a minimal list of features from the EHR datasets to ensure model performance and to enable future point-of-care risk prediction.

Results

The best-performing model uses from clinical features related to mental health history, medical comorbidity, obstetric complications, medication prescription orders, and patient demographic characteristics. The model performances as measured by area under the receiver operating characteristic curve (AUC) are 0.937 (95% CI 0.912 - 0.962) and 0.886 (95% CI 0.879-0.893) in the development and validation datasets, respectively. The model performances were consistent when tested using data ending at multiple time periods during pregnancy and at childbirth.

Limitations

The prevalence of PPD in the study data represented a treatment prevalence and is likely lower than the illness prevalence.

Conclusions

EHRs and machine learning offer the ability to identify women at risk for PPD early in their pregnancy. This may facilitate scalable and timely prevention and intervention, reducing negative outcomes and the associated burden.

Key words:

Postpartum depression, machine learning, electronic health records

1. Introduction

Postpartum depression (PPD) is a potentially life-threatening mental health condition that occurs up to one year following childbirth (1). The prevalence of PPD is estimated to affect as many as 1 in 7 mothers in the US (2, 3), but underdiagnosis and lack of treatment for PPD are common, especially among women with low socioeconomic status (4, 5). Long-term health effects of PPD to mothers, children, and family include increased maternal and infant mortality, increased hospitalizations, impaired mother-child bonding, and impaired long-term child development (6-9). The disease mechanism of PPD is multifactorial. Clinically, a history of mental illness is the most significant risk factor (1, 10). Social determinants of health (SDoH), including poor marital relationship, low socioeconomic status, and stressful life events are also known contributors to increased PPD risk (5). New research indicates that there may be additional biomarkers associated with the risk for developing PPD such as excessive proinflammatory immune system activation, possible disruptions in fatty acid metabolism, disruptions in hypothalamic-pituitary-adrenal (HPA) functioning, altered neurosteroid physiology, and genetic and epigenetic signatures (11).

The importance of PPD prevention and timely intervention cannot be overstated. The American College of Obstetricians and Gynecologists (ACOG) (12), the American Academy of Pediatrics (AAP) (13), the US Preventive Services Task Force (4), and several other organizations (1) have guidelines and recommendations for universal PPD screening as part of usual care during pregnancy and the postpartum period. Current PPD prevention strategies focus on secondary rather than primary prevention, using questionnaire-based screening instruments such as the Edinburgh Postnatal Depression Scale (EPDS) (14) and Patient Health Questionnaire-9 (PHQ-9) (15) to detect symptoms. Primary prevention techniques intervene in an illness course prior to symptom onset while secondary prevention techniques intervene soon after the symptom onset, but prior to the full manifestation of the illness. Unfortunately, it has been demonstrated that in women known to be at high risk of PPD, delaying intervention until the onset of symptoms only mildly attenuates risk for depression, while intervening with appropriately targeted prevention before the onset of symptoms substantially mitigates depression relapse risk (16). In addition to being "too little, too late" from a clinical perspective, these screening tools present major feasibility problems for both large and smaller health systems (17, 18). In order to come into compliance with current screening recommendations, obstetric practices often require substantial change, including not just changes to clinical workflows, but also staffing changes, new electronic health records (EHR) workflow builds, collaboration with

referral networks, and investment in staff and provider training. Even then, further challenges persist such as mental health-related stigma, limitations in provider time, attention, and expertise, and scarcity in specialized mental health treatment resources.

We argue that taking a primary prevention approach has the promise of reducing the investment and resources required to address PPD while at the same time reducing the incidence of PPD rates. In this work, to identify signals that may suggest elevated future risk of PPD, we propose a primary prevention approach that is data-driven, leveraging machine learning applications to EHR data (19, 20). EHR data can be collected and analyzed routinely on a large scale using machine learning, as demonstrated by successful data-driven clinical decision support (21) applications that assist with decision making across clinical conditions (22-25). We developed an end-to-end framework (Fig. 1) to extract features from EHR data for processing, including demographics, clinical diagnoses, medication prescriptions, laboratory results, and unstructured clinical notes. These data are sent to an optimization process to select important features and incorporated in multiple machine learning algorithms including regularized logistic regression, random forest, decision tree, extreme gradient boosting (XGboost), and multilayer perceptron (MLP) (26) to predict the risk of PPD. The framework was implemented and evaluated using data available at different time intervals during pregnancy (12-week, 18-week, 24-week, and 30-week) during pregnancy and after childbirth.

We aim to demonstrate that the data-driven primary intervention approach provides an opportunity for individualized therapeutic interventions such as changing screening timelines, engaging with appropriate preventive strategies, or tailoring clinician PPD counseling time according to a patient need. To the best of our knowledge, this study is among the first in developing an EHR-based machine learning framework for identifying women at risk for PPD (19, 27-29).

2. Methods

2.1 Inclusion Criteria

All pregnant women with fully completed antenatal care procedures who had live births of infants were included in the study. The exclusion criteria were (1) maternal age below 18 or above 45, or (2) lack of outpatient, inpatient or

emergency room encounter information in the EHR data within 1 year following childbirth. Participants with a prior history of mental illness and participants with active mental illness were not excluded to ensure clinical applicability in real implementation (Fig. 2). The study was approved by the Institutional Review Board at Weill Cornell Medicine (IRB protocol# 1711018789). Data extraction and analysis were performed in 2019.

2.2 Study Design

2.2.1 Outcome

The outcome is defined as having a diagnosis of PPD within 1 year of childbirth. A PPD diagnosis was defined using Systematized Nomenclature of Medicine (SNOMED) codes and the use of antidepressants within 1 year following childbirth (1, 30). The specific SNOMED codes for PPD definition are listed in Appendix (Table A1). The use of antidepressants was defined by Anatomical Therapeutic Chemical (ATC) codes under N06A (31). To ensure that antidepressants were primarily used for treatment of mental health conditions, and not for other indications such as pain, we further excluded the following medications: *Amitriptyline, Clomipramine, Duloxetine, Flupentixol, and Nortriptyline* (32).

2.2.2 Data Sources

For algorithm development, EHR data including demographics, diagnoses, medication prescriptions, procedures, laboratory measurements, and social determinants of health (SDoH) including the built environment characteristics such as distance to public transportation and green space on eligible patients were obtained at Weill Cornell Medicine (WCM) and NewYork-Presbyterian Hospital in New York City, USA between January 2015 and June 2018. For algorithm validation, EHR data was derived from multiple health systems across New York City affiliated to the Patient-Centered Outcomes Research Institute funded New York City Clinical Data Research Network data (NYC-CDRN) between August 2004 and October 2017 (33). We randomly selected 80% of the data from WCM as the training set including cross-validation and model tuning, and held the remaining 20% as the test set individually. The NYC-CDRN data was used solely as a validation set.

Both datasets were represented using Observational Medical Outcomes Partnership (OMOP) Common Data Model to record patient demographics, encounter records, diagnostic codes, procedures, prescription medications, and

laboratory measurements (34). Diagnoses, laboratory measurements, and procedures are represented as SNOMED codes, Logical Observation Identifiers Names and Codes (LOINC), and Current Procedural Terminology (CPT) codes, respectively. Medications were standardized using the ATC classification system. In addition, marital status was extracted from unstructured clinical notes using regular expression-based searches, and individuals were classified as married or not married (single/divorced/widowed) at the time of childbirth. Age was calculated as the time difference between childbirth and delivery dates. Mental health history before pregnancy was defined as having at least one diagnosis including organic disorders, substance-related disorders, schizophrenic/psychotic disorders, mood disorders, anxiety disorders, personality disorders, and other psychiatric disorders (35). Features with frequencies below 10 were omitted from the study to remove rare events during pregnancy. Mean values were used to perform the imputation of missing numerical values. Discrete features, such as clinical diagnoses, prescribed medication, were coded as dummy features (36). Numeric features were normalized in the scale of -1 to 1. Statistical comparison across the PPD and non-PPD group was performed using Stata 14. Independent sample T-test assuming unequal variances and Chi-Square test was used for continuous and categorical variables as appropriate.

2.3 Machine learning

2.3.1 Framework

The schematic diagram of our PPD prediction framework is shown in Fig. 1 (Schematic diagram of our PPD prediction framework) that describes the various steps involved in data preprocessing and risk model development. The machine learning model training was optimized using sequential forward selection (SFS) (37) – a greedy search algorithm that searches for the combination of features that returns the maximum algorithm discriminatory power (38). Starting with an empty feature set, SFS iteratively examines each feature combination such that the algorithm's performance can be maximized until the stopping criteria for the search is reached (Fig. A1) (37). Five machine learning algorithms were trained, including random forest, decision tree, extreme gradient boosting (XGboost), regularized logistic regression, and multilayer perceptron (MLP). These algorithms were developed by iteratively splitting the data available to detect collective patterns across features in the subset of the data that maximally discriminate outcome classes, followed by testing the performance on the held-out data. This training process allowed us to develop prediction algorithms that are generalizable to unseen data.

Algorithm parameters were determined using a grid search for each algorithm that comprehensively searched for the best hyperparameters and parameters that resulted in the highest model performance as measured by area under the receiver operating characteristic curve (AUC). The stopping criteria for SFS were defined as 1) no increase in the AUC by 0.001 after 10 consecutive iterations, or 2) the predetermined maximum number of feature set has been reached. SFS was performed separately for women with, and without, mental health history to ensure that the model can predict for both types of patients when in actual use. We combined features selected from both SFS into a single feature set such that a single algorithm can be used for patients with and without a history of mental illness. Using the combined features, each of the machine learning algorithms was trained using 5-fold cross-validation.

2.3.2 Expert adjudicated feature selection

Clinicians in our study team (AH and RJ) reviewed the selected features in the best performing algorithm to validate feature inclusion and ensure algorithm interpretability. Starting with the entire list of features selected by SFS, we iteratively eliminated features that were determined to be irrelevant, re-constructed the algorithm using the adjudicated features, and measured the algorithm performance. This iterative process was performed while keeping the minimum AUC at 0.8. Features that were changed during this process are listed in the Appendix (Table A2).

2.3.3 Evaluation

The evaluation was performed using the held-out data set at WCM and the entire dataset from NYC-CDRN using AUC, sensitivity, specificity, and the Brier score (39). AUC is an aggregate measure of the algorithm's ability to discriminate outcome classes across all possible classification thresholds. The Brier score measures the accuracy of prediction (40). As such, higher AUC and lower Brier score indicate better prediction performance. To evaluate the algorithm performance in a simulated gestational period where data are being accumulated during pregnancy, we computed evaluation metrics using data available up to 5 different periods. Starting with each patient's first available pregnancy encounter, we created a test dataset ending at 12-week, 18-week, 24-week, and 30-week during pregnancy, and also at childbirth assuming that data at 12-week pregnancy and childbirth contain the least and the most complete information, respectively. Lastly, error analyses were conducted by manual chart review using

patients' medical records for up to 2 years after childbirth for 150 false positives and negatives. Machine learning algorithms were trained and evaluated using Scikit-learn and Seaborn in Python (3.6.5).

3. Results

A total of 15,197 deliveries from January 2015 to June 2018 were included in our analysis, excluding 124 women below age 18 or above age 45 at the time of delivery, and 2,312 women without records of clinical encounters within 1 year following childbirth (Fig. 2). Study data were randomly split into training (N=12157) and testing (N=3040) using cross-validation. The validation set contained 53,972 deliveries from August 2004 to October 2017, after excluding 1,903 deliveries by women below age 18 or above age 45 and 15,141 deliveries without encounters recorded within 1 year after childbirth (Fig. 2). The prevalence of depression was 6.7% (N=1,010) and 6.5% (N=3,513) in the WCM and NYC-CDRN datasets, respectively.

Table 1 provides the descriptive statistics of the two datasets. We found significant differences in age, the number of emergency department (ED) visits, and racial distribution between PPD and non-PPD groups in the training and validation data, respectively. The average age at the time of delivery was 33.68 (SD=4.54) in the non-PPD group and 34.56 (SD=4.39) in the PPD group of patients in the WCM dataset (p-value<0.001), and 28.87 (SD=6.20) and 30.70 (SD=6.13) in the CDRN dataset, respectively (p-value<0.001). The number of emergency room visits in the PPD group was higher than the non-PPD group in both the WCM (1.68 ± 1.55 vs. 1.32 ± 1.24, p-value<0.001) and NYC-CDRN (6.30 ± 9.97 vs. 5.37 ± 6.87, p-value<0.001) datasets. The training and validation datasets had different distribution of PPD across racial groups. In the WCM data, the incidence rate of PPD was the highest among White women (8.8%) and the lowest among Asian women (3.0%). In the CDRN data, the rate of PPD was the highest among White women (12.43%), Black patients had the lowest rates (4.76%).

Using SFS, 32 features were selected to be incorporated in the algorithm related to patient demographic statuses, health service utilization, mental health history, newly diagnosed mental health conditions during pregnancy, other obstetric and/or medical diagnoses during pregnancy, and vital signs. As shown in Table 2, the majority (28 out of 32) of the features included in the algorithm have statistically significant association with the outcome. Features that are indicative of past and current mental health conditions and being single mothers were associated with higher

odds of a PPD diagnosis. Additionally, complications during pregnancy such as palpitations, diarrhea, vomiting, and abdominal pain also were associated with higher odds of a PPD diagnosis. Health service utilization including medication prescriptions such as Beta blocking agents, delivery by cesarean, and emergency department (ED) visits were also associated with higher odds of a PPD diagnosis. Having an Asian race was associated with lower odds of a PPD diagnosis. Fig. A2 in the Appendix shows the Pearson correlation among the features.

Evaluation results of the algorithm performance are shown in Table 3. Logistic regression with L2 regularization was found to be the best performing algorithm using data available up to childbirth. The AUC was 0.937 (95% CI: 0.912-0.962) and 0.886 (95% CI: 0.879-0.893) in the WCM and NYC-CDRN datasets, respectively. The AUC was lower in the validation dataset potentially due to the lack of certain features such as marital status which was available only in the WCM dataset. While evaluating algorithms at different periods during the pregnancy, we observed a steady performance with respect to AUC of 0.921, 0.919, 0.922, 0.921, and 0.937 using data extracted up to 12 weeks, 18 weeks, 24 weeks, 30 weeks of gestation, and at childbirth, respectively. The steady performance may be explained by the early availability and invariability of the predictive features (see Table 2). In the NYC-CDRN dataset, we observe an increase in algorithm performance as more data accumulate over time, with an AUC of 0.810, 0.817, 0.821, 0.824, and 0.886 at 12 weeks, 18 weeks, 24 weeks, 30 weeks of gestation, and lastly at childbirth, respectively. Additionally, we report positive and negative predictive values in Table 3. While negative predictive values are close to 1 for nearly all models across time periods, we find that positive predictive values are low especially in the validation site. This could be explained by the relative low prevalence of PPD and the high frequency of the patients who were not diagnosed to have PPD (based on our criteria), but were predicted so.

False-positive and false-negative results from the algorithm were evaluated by manual chart reviews of a randomly selected 150 cases that were incorrectly classified by the logistic regression classifier. The cases had 140 and 10 false positives and false negatives, respectively. PPD diagnosis after the study period and lack of proper coding were identified as two potential reasons for the false positives and negatives. For example, the manual chart review identified that 45% of the patients incorrectly predicted to develop PPD by the prediction algorithm were in fact women who were noted to be suffering from PPD in the clinical notes. Furthermore, 34% of the PPD mentions in the notes were made one year after childbirth, beyond our study period. Thus, the incorrect predictions were due to

the lack of good coding practices for PPD, a phenomenon that is frequently observed in other observational mental health studies using EHRs (41). The availability of predictors related to mental health history also presented challenges. For example, the error analysis identified the history of anxiety and depression on 36.4% of false negative cases through manual chart review. For these patients the mental health history was not coded in the structured EHR data. Extraction of features using natural language processing techniques may facilitate higher performance by the algorithm in future studies.

4. Discussion

Results from this study suggest a promising direction to leverage routinely collected EHR data to identify pregnant women at risk for PPD. Selected EHR-driven predictors characterize women's health history, pregnancy health, demographics, and healthcare utilization. Several known PPD risk factors from the literature were represented by variables extracted in the sequential feature selection process, including history of anxiety, mood disorder, and other mental disorders, antidepressant use, incidental mental health illnesses during pregnancy, cesarean section, and single motherhood.(1, 42) Our model further identifies additional comorbid predictors, including palpitations, diarrhea, vomiting during pregnancy, hypertensive disorders and hypothyroidism. Among these comorbidities, thyroid dysfunction and hypertensive disorders have been associated with PPD onset in previous literature.(43, 44) Palpitation, a common cardiac symptom, may also be a symptom of depression that was discovered by the model.(45, 46) In addition, medication prescriptions of beta blocking agents and antihistamines were identified as predictors. Literature has reported the use of both beta blockers and antihistamines in association with depression although not conclusively (47-49). Related to mode of delivery, our model selected cesarean section as a risk factor for PPD, as also studied in the previous literature.(50, 51) Lastly, the number of ED visits during pregnancy and postpartum may be an indicator of a lack of proper access to primary and obstetric care.(52)

As seen in our experiments, the risk computed by the PPD prediction algorithm updates in response to the new health information that accumulates overtime with repeated visits during pregnancy, thereby potentially allowing care providers to take timely actions according to the risk evolution. (12, 13) With these automatically extractable features, an EHR-based prediction tool may assist with existing EHR interventions for screening to minimize variations across clinical practices in screening and information collection.(53) Previous studies have reported that

while the rates of screening and referrals for mental health care can be high when obstetricians recognize a risk for PPD, but they are low if symptoms are unnoticed by the care provider.(54) Our risk prediction model, by identifying women with elevated risk, may assist with tacitly raising clinician awareness of PPD and potentially increasing screening and referral rates.

5. Limitations

Several limitations exist in our research. First, our study cohort as derived from the EHR in an urban academic medical center is not representative of the general US population suffering from PPD and differs from cohorts reported in previous studies with respect to PPD prevalence (2). This prevalence is likely the treatment prevalence rather than the illness prevalence, as the data may not capture patients outside of the studied health system and geographical location. The prevalence may also reflect the clinician coding practices on recording a diagnosis of PPD at the study sites. Persistent stigma and social consequences of having depression coded in the EHR may prevent providers from 'officially' coding the diagnosis even if it is made clinically. Further, also due to stigma, patients may withhold symptom information from providers preventing accurate diagnosis. In addition to using diagnostic codes, we also defined PPD using antidepressant use while excluding those for pain indications. However, it is possible that some antidepressants were used for anxiety rather than PPD. Anxiety disorders are so frequently comorbid with depression in the peripartum period such that a diagnosis of one may even be a proxy for unidentified depression. Thus, we decided it was important not to exclude anxiety disorder indications even at the expense of specificity, although we recognize this as a limitation of our study. Our ongoing and future work will attempt to parse these indications further by applying natural language processing to the unstructured clinical notes.

Relatedly, in this study, we did not specifically include only patients with incident depression. This decision was meant to acknowledge the powerful effect that mental health histories have on risk for developing PPD as well as to provide a clinically meaningful risk stratification for real-world obstetric providers who have large cohorts of patients with mental health histories and those who are actively seeking treatment in their practices. Due to the lack of comprehensive screening at our health systems and clinics in the study sites, we did not capture EPDS and PHQ-9 scores to define PPD. We also did not compare effectiveness of primary prevention via the prediction algorithm to current widely recommended secondary prevention efforts via EPDS or PHQ-9 screening. However, we did

compare with algorithms reported in prior literature as a potential primary intervention approach, and demonstrated improved model performance. Compared to prior work by Camdeviren et al (55), Tortajada et al (27), and Natarajan et al (56), our algorithm was built by exhaustively selecting most predictive features from a larger number of candidate features from the EHR data, with an eventual goal of integrating such risk prediction models within the EHR systems and clinical workflows. Furthermore, compared to our initial pilot work (19) which did not include prior mental health diagnosis and treatment history as predictors, the prediction algorithm from this study demonstrated a significant increase in AUC, sensitivity, and specificity.

A number of future works are under preparation to address these limitations. We found White and Asian races to be predictive features in this study. However, a substantial proportion of race was unknown in both the training and validation datasets, potentially due to lack of proper documentation in the EHR (57). This is an important area for further consideration in future studies.(58) These include a comparison of the data-driven primary intervention against usual care as a clinical trial, and additional validation work at study sites in the greater US and abroad using datasets with different PPD prevalence to evaluate the algorithm generalizability. While findings from this study present a promise for PPD risk identification using available EHR data, we realize that EHR data capture only a limited portion of patients' lives which contribute to PPD. Therefore, we will also evaluate whether the addition of patient-reported outcomes or information derived from mobile health devices, such as wearables, can contribute to higher algorithm performance. Lastly, improvement in the machine learning framework will include techniques to adjust for differing outcome distributions such that the method can be more generally applied to other populations.

6. Conclusions

In summary, this study demonstrates that a data-driven primary intervention approach using machine learning and EHR data may be leveraged to reduce the healthcare provider burden of identifying PPD risk. Methods created in this study may pave a path towards data-driven, accurate, and scalable clinical decision support for PPD risk identification with potential benefits through early prevention, diagnosis, and intervention.

References

Alijaniha, F., Noorbala, A., Afsharypuor, S., Naseri, M., Fallahi, F., Mosaddegh, M., Zadeh, S.F., Sadrai, S., 2016. Relationship Between Palpitation and Mental Health. Iran Red Crescent Me 18.
Barsky, A.J., Cleary, P.D., Coeytaux, R.R., Ruskin, J.N., 1994. Psychiatric disorders in medical outpatients complaining of palpitations. J Gen Intern Med 9, 306-313.
Beck, C.T., Gable, R.K., 2000. Postpartum Depression Screening Scale: development and psychometric testing. Nurs Res 49, 272-282.
Biaggi, A., Conroy, S., Pawlby, S., Pariante, C.M., 2016. Identifying the women at risk of antenatal anxiety and depression: A systematic review. J Affect Disord 191, 62-77.
Bishop, C.M., 2006. Pattern recognition and machine learning. Springer-Verlag New York.
Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 30, 1145-1159.
Camdeviren, H.A., Yazici, A.C., Akkus, Z., Bugdayci, R., Sungur, M.A., 2007. Comparison of logistic regression model and classification tree: An application to postpartum depression data. Expert Systems with Applications 32, 987-994.
Canada, S., 2015. Mental disorders, conditions, and ICD-9 and ICD-10-CA codes. Statistics Canada.
Carter, F.A., Frampton, C.M.A., Mulder, R.T., 2006. Cesarean section and postpartum depression: A review of the evidence examining the link. Psychosom Med 68, 321-330.
Cohen, L.S., Altshuler, L.L., Harlow, B.L., Nonacs, R., Newport, D.J., Viguera, A.C., Suri, R., Burt, V.K., Hendrick, V., Reminick, A.M., Loughead, A., Vitonis, A.F., Stowe, Z.N., 2006. Relapse of major depression during pregnancy in women who maintain or discontinue antidepressant treatment. Jama 295, 499-507.
Committee, A., 2018. ACOG Committee Opinion No. 757: Screening for Perinatal Depression. Obstet Gynecol 132, e208-e212.
Cox, J.L., Holden, J.M., Sagovsky, R., 1987. Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 150, 782-786.
Dietz, P.M., Williams, S.B., Callaghan, W.M., Bachman, D.J., Whitlock, E.P., Hornbrook, M.C., 2007. Clinically identified maternal depression before, during, and after pregnancies ending in live births. Am J Psychiatry 164, 1515-1520.
Earls, M.F., Yogman, M.W., Mattson, G., Rafferty, J., 2019. Incorporating Recognition and Management of Perinatal Depression Into Pediatric Practice. Pediatrics 143.
Field, T., 2010. Postpartum depression effects on early interactions, parenting, and safety practices: a review. Infant Behav Dev 33, 1-6.
Forman, D.N., Videbech, P., Hedegaard, M., Salvig, J.D., Secher, N.J., 2000. Postpartum depression: identification of women at risk. Brit J Obstet Gynaec 107, 1210-1217.
Gerstman, B.B., Jolson, H.M., Bauer, M., Cho, P., Livingston, J.M., Platt, R., 1996. The incidence of depression in new users of beta-blockers and selected antihypertensives. J Clin Epidemiol 49, 809-815.
Gjerdingen, D.K., Yawn, B.P., 2007. Postpartum depression screening: importance, methods, barriers, and recommendations for practice. J Am Board Fam Med 20, 280-288.
Goldstein, B.A., Navar, A.M., Pencina, M.J., Ioannidis, J.P., 2017. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 24, 198-208.
Goodman, J.H., Tyer-Viola, L., 2010. Detection, Treatment, and Referral of Perinatal Depression and Anxiety by Obstetrical Providers. J Womens Health 19, 477-490.

Hahn-Holbrook, J., Cornwell-Hinrichs, T., Anaya, I., 2017. Economic and Health Predictors of National Postpartum Depression Prevalence: A Systematic Review, Meta-analysis, and Meta-Regression of 291 Studies from 56 Countries. Front Psychiatry 8, 248.
Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29-36.
Jacques, N., de Mola, C.L., Joseph, G., Mesenburg, M.A., da Silveira, M.F., 2019. Prenatal and postnatal maternal depression and infant hospitalization and mortality in the first year of life: A systematic review and meta-analysis. J Affect Disord 243, 201-208.
Jiménez-Serrano, S., Tortajada, S., García-Gómez, J.M., 2015. A Mobile Health Application to Predict Postpartum Depression Based on Machine Learning. Telemed J E Health 21, 567-574.
Kaushal, R., Hripcsak, G., Ascheim, D.D., Bloom, T., Campion, T.R., Jr., Caplan, A.L., Currie, B.P., Check, T., Deland, E.L., Gourevitch, M.N., Hart, R., Horowitz, C.R., Kastenbaum, I., Levin, A.A., Low, A.F., Meissner, P., Mirhaji, P., Pincus, H.A., Scaglione, C., Shelley, D., Tobin, J.N., 2014. Changing the research landscape: the New York City Clinical Data Research Network. J Am Med Inform Assoc 21, 587-590.
Le Donne, M., Mento, C., Settineri, S., Antonelli, A., Benvenga, S., 2017. Postpartum Mood Disorders and Thyroid Autoimmunity. Front Endocrinol (Lausanne) 8, 91.
Lee, W.C., Veeranki, S.P., Serag, H., Eschbach, K., Smith, K.D., 2016. Improving the Collection of Race, Ethnicity, and Language Data to Reduce Healthcare Disparities: A Case Study from an Academic Medical Center. Perspect Health Inf Manag 13, 1g.
Liang, H., Tsui, B.Y., Ni, H., Valentim, C.C.S., Baxter, S.L., Liu, G., Cai, W., Kermany, D.S., Sun, X., Chen, J., He, L., Zhu, J., Tian, P., Shao, H., Zheng, L., Hou, R., Hewett, S., Li, G., Liang, P., Zang, X., Zhang, Z., Pan, L., Cai, H., Ling, R., Li, S., Cui, Y., Tang, S., Ye, H., Huang, X., He, W., Liang, W., Zhang, Q., Jiang, J., Yu, W., Gao, J., Ou, W., Deng, Y., Hou, Q., Wang, B., Yao, C., Liang, Y., Zhang, S., Duan, Y., Zhang, R., Gibson, S., Zhang, C.L., Li, O., Zhang, E.D., Karin, G., Nguyen, N., Wu, X., Wen, C., Xu, J., Xu, W., Wang, B., Wang, W., Li, J., Pizzato, B., Bao, C., Xiang, D., He, W., He, S., Zhou, Y., Haw, W., Goldbaum, M., Tremoulet, A., Hsu, C.N., Carter, H., Zhu, L., Zhang, K., Xia, H., 2019. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 25, 433-438.
Long, M.M., Cramer, R.J., Jenkins, J., Bennington, L., Paulson, J.F., 2019. A systematic review of interventions for healthcare professionals to improve screening and referral for perinatal mood and anxiety disorders. Arch Women Ment Hlth 22, 25-36.
Loudon, H., Nentin, F., Silverman, M.E., 2016. Using clinical decision support as a means of implementing a universal postpartum depression screening program. Arch Womens Ment Health 19, 501-505.
Löwe, B., Unützer, J., Callahan, C.M., Perkins, A.J., Kroenke, K., 2004. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 42, 1194-1201.
Meltzer-Brody, S., Howard, L.M., Bergink, V., Vigod, S., Jones, I., Munk-Olsen, T., Honikman, S., Milgrom, J., 2018. Postpartum psychiatric disorders. Nat Rev Dis Primers 4, 18022.
O'Connor, E., Senger, C.A., Henninger, M.L., Coppola, E., Gaynes, B.N., 2019. Interventions to Prevent Perinatal Depression: Evidence Report and Systematic Review for the US Preventive Services Task Force. Jama 321, 588-601.
Overhage, J.M., Ryan, P.B., Reich, C.G., Hartzema, A.G., Stang, P.E., 2012. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc 19, 54-60.

Ozdemir, P.G., Karadag, A.S., Selvi, Y., Boysan, M., Bilgili, S.G., Aydin, A., Onder, S., 2014. Assessment of the effects of antihistamine drugs on mood, sleep quality, sleepiness, and dream anxiety. Int J Psychiat Clin 18, 161-168.
Petersen, I., Peltola, T., Kaski, S., Walters, K.R., Hardoon, S., 2018. Depression, depressive symptoms and treatments in women who have recently given birth: UK cohort study. BMJ Open 8, e022152.
Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Liu, P.J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G.E., Irvine, J., Le, Q., Litsch, K., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboum, S.L., Chou, K., Pearson, M., Madabushi, S., Shah, N.H., Butte, A.J., Howell, M.D., Cui, C., Corrado, G.S., Dean, J., 2018. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 1, 18.
Rodríguez, P., Bautista, M.A., Gonzàlez, J., Escalera, S., 2018. Beyond one-hot encoding: Lower dimensional target embedding. Image and Vision Computing 75, 21-31.
Rufibach, K., 2010. Use of Brier score to assess binary predictions. J Clin Epidemiol 63, 938-939; author reply 939.
S., N., A., P., N., R., A., B., K., S., K., C., 2017. Boosting for Postpartum Depression Prediction, 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), pp. 232-240.
Schofield, P., Das-Munshi, J., Mathur, R., Congdon, P., Hull, S., 2016. Does depression diagnosis and antidepressant prescribing vary by location? Analysis of ethnic density associations using a large primary-care dataset. Psychol Med 46, 1321-1329.
Serati, M., Redaelli, M., Buoli, M., Altamura, A.C., 2016. Perinatal Major Depression Biomarkers: A systematic review. J Affect Disord 193, 391-404.
Sheen, J.J., Smith, H.A., Tu, B., Liu, Y., Sutton, D., Bernstein, P.S., 2019. Risk Factors for Postpartum Emergency Department Visits in an Urban Population. Matern Child Hlth J 23, 557-566.
Sholle, E.T., Pinheiro, L.C., Adekkanattu, P., Davila, M.A., Johnson, S.B., Pathak, J., Sinha, S., Li, C., Lubansky, S.A., Safford, M.M., Campion, T.R., 2019. Underserved populations with missing race ethnicity data differ significantly from those with structured race/ethnicity documentation. Journal of the American Medical Informatics Association : JAMIA 26, 722-729.
Shortliffe, E.H., Sepúlveda, M.J., 2018. Clinical Decision Support in the Era of Artificial Intelligence. Jama 320, 2199-2200.
Stein, A., Pearson, R.M., Goodman, S.H., Rapa, E., Rahman, A., McCallum, M., Howard, L.M., Pariante, C.M., 2014. Effects of perinatal mental disorders on the fetus and child. Lancet 384, 1800-1819.
Stewart, C.C., Lu, C.Y., Yoon, T.K., Coleman, K.J., Crawford, P.M., Lakoma, M.D., Simon, G.E., 2019. Impact of ICD-10-CM Transition on Mental Health Diagnoses Recording. EGEMS (Wash DC) 7, 14.
Stewart, D.E., Vigod, S., 2016. Postpartum Depression. N Engl J Med 375, 2177-2186.
Strapasson, M.R., Ferreira, C.F., Ramos, J.G.L., 2018. Associations between postpartum depression and hypertensive disorders of pregnancy. Int J Gynecol Obstet 143, 367-373.
T., L., G., T., 2015. SFS feature selection technique for multistage emotion recognition, 2015 IEEE 3rd Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), pp. 1-4.

Tomašev, N., Glorot, X., Rae, J.W., Zielinski, M., Askham, H., Saraiva, A., Mottram, A., Meyer, C., Ravuri, S., Protsyuk, I., Connell, A., Hughes, C.O., Karthikesalingam, A., Cornebise, J., Montgomery, H., Rees, G., Laing, C., Baker, C.R., Peterson, K., Reeves, R., Hassabis, D., King, D., Suleyman, M., Back, T., Nielson, C., Ledsam, J.R., Mohamed, S., 2019. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 572, 116-119.
Tortajada, S., García-Gomez, J.M., Vicente, J., Sanjuán, J., de Frutos, R., Martín-Santos, R., García-Esteve, L., Gornemann, I., Gutiérrez-Zotes, A., Canellas, F., Carracedo, A., Gratacos, M., Guillamat, R., Baca-García, E., Robles, M., 2009. Prediction of postpartum depression using multilayer perceptrons and pruning. Methods Inf Med 48, 291-298.
Wang, S., Pathak, J., Zhang, Y., 2019. Using Electronic Health Records and Machine Learning to Predict Postpartum Depression. Stud Health Technol Inform 264, 888-892.
Weobong, B., ten Asbroek, A.H., Soremekun, S., Gram, L., Amenga-Etego, S., Danso, S., Owusu-Agyei, S., Prince, M., Kirkwood, B.R., 2015. Association between probable postnatal depression and increased infant mortality and morbidity: findings from the DON population-based cohort study in rural Ghana. BMJ Open 5, e006509.
Wisner, K.L., Sit, D.K., McShea, M.C., Rizzo, D.M., Zoretich, R.A., Hughes, C.L., Eng, H.F., Luther, J.F., Wisniewski, S.R., Costantino, M.L., Confer, A.L., Moses-Kolko, E.L., Famy, C.S., Hanusa, B.H., 2013. Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. JAMA Psychiatry 70, 490-498.
Xu, H., Ding, Y., Ma, Y., Xin, X.L., Zhang, D.F., 2017. Cesarean section and risk of postpartum depression: A meta-analysis. J Psychosom Res 97, 118-126.
Yudofsky, S.C., 1992. Beta-blockers and depression. The clinician's dilemma. JAMA : the journal of the American Medical Association 267, 1826-1827.
Zhang, W., Liu, H., Silenzio, V.M.B., Qiu, P., Gong, W., 2020. Machine Learning Models for the Prediction of Postpartum Depression: Application and Comparison Based on a Cohort Study. JMIR Med Inform 8, e15516.

Fig. 1. Schematic diagram of our PPD prediction framework. This diagram depicts the process of the study. The modified version of multilayer perceptron (gray) is used to compare with Tortajada et al (2009).

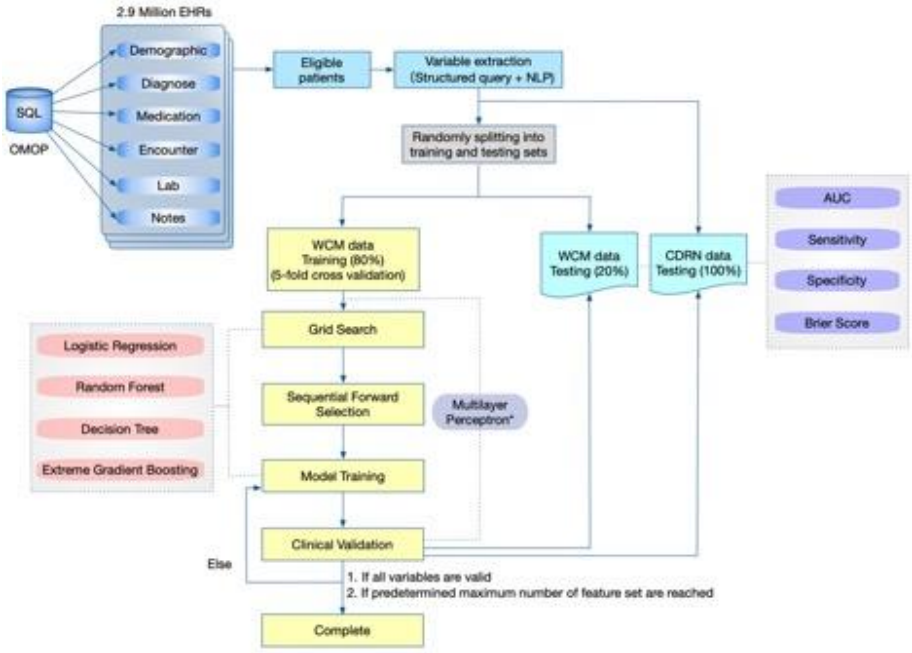


Fig. 2. The inclusion and exclusion criteria of this study

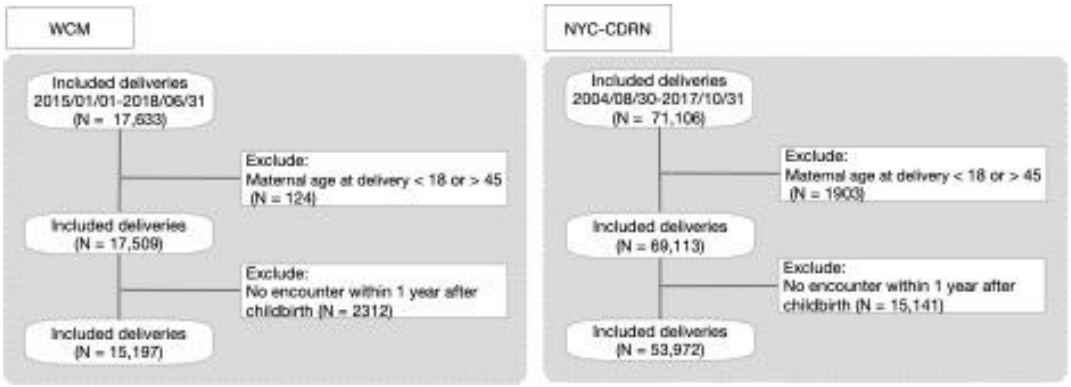


Table 1. Demographic information of the two datasets

Variable	WCM data			CDRN data		
	non-PPD	PPD	*P*	non-PPD	PPD	*P*
N (%)	14187(93.35)	1010(6.65)		50459(93.49)	3513(6.51)	
	Mean ± SD			Mean ± SD		
Age, year	33.68 ± 4.54	34.56 ± 4.39	<0.001	28.87 ± 6.20	30.70 ± 6.13	<0.001
Pre-pregnancy BMI, kg/m^2	23.92 ± 4.32	24.07 ± 4.51	0.294	28.72 ± 6.29	28.27 ± 6.91	<0.001
# ED visits	1.32 ± 1.24	1.68 ± 1.55	<0.001	5.37 ± 6.87	6.30 ± 9.97	<0.001
	N (%)			N (%)		
Race group						
White	6794(91.23)	653(8.77)	<0.001	8275(87.57)	1175(12.43)	<0.001
Asian	2784(96.97)	87(3.03)		1535(94.58)	88(5.42)	
Black	839(93.85)	55(6.15)		13815(95.24)	690(4.76)	
Other combinations	1612(94.16)	100(5.84)		19163(94.09)	1204(5.91)	
Unknown	2158(94.94)	115(5.06)		7671(95.56)	356(4.44)	

Table 2. The association of selected features with PPD using univariate logistic regression in WCM training data.

Variables	non-PPD	PPD	OR (95%CI)	*P*
N(%)	11324(93.15)	833(6.85)		
Anxiety history	135(1.19)	473(56.78)	108.90(87.55,135.44)	<0.001
Other disorder history	156(1.38)	169(20.29)	18.22(14.46,22.96)	<0.001
Antidepressants	22(0.19)	129(15.49)	94.14(59.52,148.89)	<0.001
Mood disorder history	120(1.06)	266(31.93)	43.80(34.75,55.21)	<0.001
Depression in pregnancy	60(0.53)	108(12.97)	27.97(20.22,38.68)	<0.001
Anxiety in pregnancy	82(0.72)	141(16.93)	27.93(21.05,37.07)	<0.001
Mental disorder in pregnancy	62(0.55)	91(10.92)	22.27(15.99,31.01)	<0.001
Palpitations	147(1.30)	28(3.36)	2.65(1.75,3.99)	<0.001
Diarrhea	159(1.40)	29(3.48)	2.54(1.70,3.79)	<0.001
Vomiting in pregnancy	298(2.63)	53(6.36)	2.52(1.86,3.40)	<0.001
Hypertensive disorder	104(0.92)	17(2.04)	2.25(1.34,3.77)	0.002
Acute pharyngitis	107(0.94)	17(2.04)	2.19(1.30,3.66)	0.003
Hemorrhage in early pregnancy antepartum	92(0.81)	14(1.68)	2.09(1.18,3.68)	0.011
White	5390(47.60)	540(64.83)	2.03(1.75,2.35)	<0.001
Threatened miscarriage	478(4.22)	67(8.04)	1.99(1.52,2.59)	<0.001
Abdominal pain	699(6.17)	96(11.52)	1.98(1.58,2.48)	<0.001
Migraine	69(0.61)	10(1.20)	1.98(1.02,3.86)	0.044
Beta blocking agents	173(1.53)	24(2.88)	1.91(1.24,2.95)	0.003
Antihistamines for systemic use	661(5.84)	84(10.08)	1.81(1.43,2.30)	<0.001
Hypothyroidism	1062(9.38)	121(14.53)	1.64(1.34,2.01)	<0.001
Placental infarct	264(2.33)	31(3.72)	1.62(1.11,2.37)	0.013
Single (vs. Married)	1412(12.47)	154(18.49)	1.59(1.33,1.91)	<0.001
Deliveries by cesarean	2449(21.63)	240(28.81)	1.47(1.25,1.72)	<0.001
Direct acting antivirals	482(4.26)	50(6.00)	1.44(1.06,1.94)	0.018
Primigravida	6699(59.16)	556(66.75)	1.39(1.19,1.61)	<0.001
Pre-eclampsia	79(0.70)	8(0.96)	1.38(0.67,2.87)	0.386
Other antibacterials	489(4.32)	47(5.64)	1.33(0.97,1.80)	0.073

#ED visit	1.31 ± 1.24	1.67 ± 1.54	1.24(1.17,1.31)	<0.001
Abnormality of organs and/or soft tissues of pelvis affecting pregnancy	659(5.82)	53(6.36)	1.10(0.82,1.47)	0.52
Diastolic blood pressure in third trimester	69.30 ± 5.64	69.83 ± 5.88	1.09(1.02,1.16)	0.009
False labor at or after 37 completed weeks of gestation	270(2.38)	13(1.56)	0.65(0.37,1.14)	0.131
Asian	2227(19.67)	75(9.00)	0.40(0.32,0.52)	<0.001

Table 3. The model performance.

Time	Classifier	AUROC	Sensitivity	Specificity	Brier Score	PPV	NPV
		WCM (development site)					
12wk	Logistics regression	0.921(0.893,0.949)	0.79	0.97	0.074	0.61	0.99
	Random Forest	0.897(0.866,0.928)	0.80	0.97	0.054	0.60	0.99
	Decision Tree	0.903(0.873,0.933)	0.83	0.96	0.045	0.59	0.99
	XGboost	0.908(0.878,0.938)	0.82	0.97	0.068	0.60	0.99
	MLP	0.921(0.893,0.949)	0.63	0.98	0.028	0.71	0.98
18wk	Logistics regression	0.919(0.891,0.947)	0.79	0.97	0.074	0.61	0.99
	Random Forest	0.897(0.866,0.928)	0.80	0.97	0.056	0.60	0.99
	Decision Tree	0.890(0.858,0.922)	0.82	0.96	0.048	0.59	0.99
	XGboost	0.902(0.872,0.932)	0.82	0.97	0.097	0.60	0.99
	MLP	0.919(0.891,0.947)	0.63	0.98	0.028	0.71	0.98
24wk	Logistics regression	0.922(0.895,0.949)	0.79	0.97	0.074	0.61	0.99
	Random Forest	0.903(0.873,0.933)	0.80	0.97	0.057	0.60	0.99
	Decision Tree	0.895(0.864,0.926)	0.83	0.96	0.048	0.59	0.99
	XGboost	0.919(0.891,0.947)	0.83	0.96	0.082	0.57	0.99
	MLP	0.920(0.892,0.948)	0.63	0.98	0.028	0.72	0.98
30wk	Logistics regression	0.921(0.893,0.949)	0.79	0.97	0.074	0.61	0.99
	Random Forest	0.914(0.885,0.943)	0.83	0.97	0.056	0.65	0.99
	Decision Tree	0.887(0.855,0.919)	0.82	0.96	0.048	0.59	0.99
	XGboost	0.912(0.883,0.941)	0.82	0.96	0.085	0.57	0.99
	MLP	0.917(0.889,0.945)	0.64	0.98	0.028	0.72	0.98
Childbirth	Logistics regression	0.937(0.912,0.962)	0.83	0.96	0.082	0.59	0.99
	Random Forest	0.935(0.910,0.960)	0.84	0.96	0.067	0.57	0.99
	Decision Tree	0.911(0.882,0.940)	0.87	0.96	0.052	0.55	0.99
	XGboost	0.935(0.910,0.960)	0.87	0.94	0.101	0.46	0.99
	MLP	0.933(0.907,0.959)	0.64	0.99	0.026	0.75	0.98
		CDRN (validation site)					
12wk	Logistics regression	0.810(0.801,0.819)	0.70	0.85	0.150	0.24	0.98
	Random Forest	0.788(0.779,0.797)	0.71	0.85	0.144	0.24	0.98
	Decision Tree	0.790(0.781,0.799)	0.71	0.85	0.152	0.24	0.71
	XGboost	0.789(0.780,0.798)	0.71	0.85	0.180	0.24	0.98
	MLP	0.812(0.803,0.821)	0.65	0.87	0.111	0.26	0.97
18wk	Logistics regression	0.817(0.808,0.826)	0.70	0.85	0.151	0.24	0.98
	Random Forest	0.794(0.785,0.803)	0.72	0.84	0.145	0.24	0.98
	Decision Tree	0.794(0.785,0.803)	0.72	0.84	0.152	0.24	0.98
	XGboost	0.793(0.784,0.802)	0.72	0.85	0.180	0.25	0.98
	MLP	0.817(0.808,0.826)	0.65	0.87	0.111	0.26	0.97

24wk	Logistics regression	0.821(0.812,0.830)	0.71	0.85	0.152	0.25	0.98
	Random Forest	0.800(0.791,0.809)	0.73	0.84	0.146	0.24	0.98
	Decision Tree	0.799(0.790,0.808)	0.73	0.84	0.152	0.24	0.98
	XGboost	0.798(0.789,0.807)	0.73	0.85	0.180	0.25	0.98
	MLP	0.824(0.815,0.833)	0.64	0.88	0.110	0.27	0.97
30wk	Logistics regression	0.824(0.815,0.833)	0.72	0.85	0.153	0.24	0.98
	Random Forest	0.807(0.798,0.816)	0.74	0.84	0.148	0.24	0.98
	Decision Tree	0.802(0.793,0.811)	0.73	0.84	0.152	0.24	0.98
	XGboost	0.801(0.792,0.810)	0.73	0.84	0.181	0.25	0.98
	MLP	0.827(0.818,0.836)	0.65	0.88	0.110	0.27	0.97
Childbirth	Logistics regression	0.886(0.879,0.893)	0.80	0.84	0.158	0.26	0.98
	Random Forest	0.860(0.852,0.868)	0.82	0.87	0.154	0.26	0.99
	Decision Tree	0.856(0.848,0.864)	0.86	0.84	0.149	0.27	0.99
	XGboost	0.864(0.856,0.872)	0.84	0.84	0.178	0.27	0.99
	MLP	0.887(0.880,0.894)	0.66	0.88	0.105	0.28	0.97

Supplemental Materials

Table S1. The association of selected features with PPD using univariate logistic regression in CDRN training data.

Variables	non-PPD	PPD	OR (95%CI)	*P*
N(%)	52091(94.21)	3199(5.79)		
Anxiety history	3280(6.30)	1242(38.82)	9.44(8.72,10.22)	<0.001
Other disorder history	3038(5.83)	761(23.79)	5.04(4.61,5.51)	<0.001
Antidepressants	309(0.59)	330(10.32)	19.28(16.43,22.62)	<0.001
Mood disorder history	4462(8.57)	1441(45.05)	8.75(8.11,9.44)	<0.001
Depression in pregnancy	2043(3.92)	755(23.60)	7.57(6.89,8.30)	<0.001
Anxiety in pregnancy	430(0.83)	1170(36.57)	69.28(61.55,78.11)	<0.001
Mental disorder in pregnancy	3110(5.97)	1523(47.61)	14.31(13.23,15.48)	<0.001
Palpitations	808(1.55)	118(3.69)	2.43(1.99,2.95)	<0.001
Diarrhea	1118(2.15)	126(3.94)	1.87(1.54,2.25)	<0.001
Vomiting in pregnancy	1749(3.36)	190(5.94)	1.82(1.55,2.11)	<0.001
Acute pharyngitis	1050(2.02)	90(2.81)	1.41(1.12,1.74)	0.002
Hemorrhage in early pregnancy antepartum	2086(4.00)	136(4.25)	1.06(0.89,1.27)	0.490
White	9229(17.72)	1093(34.17)	2.41(2.23,2.60)	<0.001
Threatened miscarriage	4165(8.00)	251(7.85)	0.98(0.86,1.12)	0.762
Abdominal pain	9750(18.72)	696(21.76)	1.21(1.11,1.32)	<0.001
Migraine	1568(3.01)	209(6.53)	2.25(1.94,2.61)	<0.001
Beta blocking agents	136(0.26)	90(2.81)	11.06(8.43,14.44)	<0.001
Antihistamines for systemic use	730(1.40)	288(9.00)	6.96(6.04,8.01)	<0.001
Hypothyroidism	1386(2.66)	245(7.66)	3.03(2.63,3.49)	<0.001
Deliveries by cesarean	73(0.14)	21(0.66)	4.71(2.82,7.52)	<0.001
Direct acting antivirals	553(1.06)	189(5.91)	5.85(4.93,6.92)	<0.001
Primigravida	19016 (36.51)	1148(35.89)	0.97(0.90,1.05)	0.480
Pre-eclampsia	1129(2.17)	86(2.69)	1.25(0.99,1.55)	0.052
Other antibacterials	994(1.91)	288(9.00)	5.09(4.43,5.82)	<0.001

#ED visit	4.79 ± 6.42	5.70 ± 9.50	1.01(1.01,1.02)	<0.001
Abnormality of organs and/or soft tissues of pelvis affecting pregnancy	1708(3.28)	133(4.16)	1.28(1.06,1.53)	0.007
Diastolic blood pressure in third trimester	109.96 ± 0.10	109.96 ± 0.03	1.05(0.74,1.37)	0.754
False labor at or after 37 completed weeks of gestation	32(0.06)	37(1.16)	19.04(11.85,30.75)	<0.001
Asian	1586(3.04)	86(2.69)	0.88(0.70,1.09)	0.254

Aim 1.2 Identifying Urban Built Environment Factors in Pregnancy Care and Maternal Mental Health Outcomes

Abstract

Backgrounds: Environmental risk factors related to the built environment have been associated with women's mental health and preventive care. This study sought to identify built environment factors that are associated with variations in prenatal care and subsequent pregnancy-related outcomes in an urban setting.

Methods: In a retrospective observational study using machine learning, we characterized the types and frequency of events in prenatal care that are associated with the various built environment factors of the patients' residing neighborhoods. We hypothesize that, in comparison to women living in high-quality built environments, women who reside in low-quality built environments experience a different pattern of clinical events that may increase the risk for adverse outcomes. Using machine learning, we performed pattern detection to characterize the variability in prenatal care with respect to encounter types, clinical problems, and medication prescriptions. Structural equation modeling was used to test the associations among built environment, prenatal care variation, and pregnancy outcome. The main outcome is postpartum depression (PPD) diagnosis within 1 year following childbirth. The exposures were the quality of the built environment in the patients' residing neighborhoods. Electronic health records (EHR) data of pregnant women (n=8,949) who had live delivery at an urban academic medical center in 2015 to 2017 were included in the study.

Results: We discovered prenatal care patterns that were summarized into three common types. Women who experienced the prenatal care pattern with the highest rates of PPD were more likely to reside in neighborhoods with homogeneous land use, lower walkability, lower air pollutant concentration, and lower accessibility to retail stores after adjusting for age, neighborhood average education level, marital status, and income inequality.

Conclusions: In an urban setting, multi-purpose and walkable communities were found to be associated with a lower risk of PPD. Findings may inform urban design policies and provide awareness for care providers on the association of patients' residing neighborhoods and healthy pregnancy.

Keywords

Pregnancy care, postpartum depression, built environment

Background

The built environment, referring to the surroundings and physical artifacts of where humans live, is considered to be one of the five major social determinants of health (SDoH).(59) The built environment determines housing quality, mode of transportation, and exposure to pollutants, effectively influencing our way of life.(60) Poor built environment causes adverse effects on physical and mental health by disrupting sleep, hindering healthy life styles, and lowering access to healthcare.(61-63) There is a gender difference on the association between the built environment and health. Mullings et al. reported an increased risk of depression among female associated with living in an unplanned neighborhood characterized by inadequate sewer treatment, water supply, and dependable supply of electricity.(64) Furthermore, the Chicago Community Adult Health Study found the women's use of preventive care to be associated with objective and perceived neighborhood support and stressors such as odors, presence of trees, and noise levels.(65)

The existing literature motivated this study to examine the impact of the built environment on health and healthcare utilization among women, and particularly, the pregnant population.(66-68) Levels of prenatal care vary across the United States.(69-71) A substantial proportion of pregnant women, in particular those with a higher comorbidity burden or low health literacy, seek and depend on care provided by emergency departments (ED) rather than primary and obstetric care.(71-73) The lack of adequate prenatal care is considered to be a risk factor for poor pregnancy outcomes and lack of proper postpartum care for mothers and infants.(74) Previous studies have studied the built environment on maternal health and birth outcomes including birth weight, gestational age, Apgar score, and newborn intensive care unit admission rates.(63, 75) Yet, evidence is still accumulating on how the built environment affects the variability in prenatal care and maternal mental health outcomes. In particular, few studied the concurrent impacts of prenatal care and built environment on mental health outcomes. Existing studies have commonly relied on the subjective perceived measures obtained from interviews and questionnaires.(62, 65, 76)

However, relying on subjective measurements may increase recall bias which occurs when some participants recall the exposure differently than others.

In this study, we hypothesize that the built environment, through influencing the accessibility to transportation, green space, safe neighborhood, and other urban structure, is associated with variability in prenatal care and subsequent maternal mental health outcomes. Given findings from previous literature on the impact of the built environment on women's mental health and use of healthcare, we defined postpartum depression (PPD) as our primary outcome.(2) PPD has been associated with increased infant mortality, higher rates of hospitalizations, impaired mother-child attachment, developmental problems in children, and increased stress within families.(77-80) The plethora of physical and psychological effects of PPD reported in previous studies include postpartum weight retention, reduced physical health, bodily pain, anxiety, low self-esteem, risky addictive behavior of substances, and suicide ideation.(81) The biological risk factors of PPD include genetic factors, age, pregnancy complications, medical illness, and smoking during pregnancy.(62, 82-84) The social, cultural, and environmental risk factors include income status, domestic violence, lack of social support, quantity and quality of green spaces, and residential noise pollution.(83, 85-89)

We tested our hypotheses by linking patients' health data extracted from de-identified electronic health records (EHRs) with publicly available census-tract level data on the built environment. Routinely collected from clinical encounters, EHR data capture detailed longitudinal health data on health and health service utilizations. Increasingly, EHR data have been used as a source of longitudinal data in population health studies for its ability to provide detailed and rich health information within patient cohorts.(90) Leveraging a large cohort of nearly 9,000 women in New York City from 2015 to 2017, we applied machine learning algorithms to EHR data to identify patterns in prenatal care.(91) We then evaluated the relationships among prenatal care patterns, PPD incidence, and the built environment using structural equation modeling.(92) The association found may inform patients, care providers, and public health policy makers in supporting healthy pregnancy and new motherhood.

Methods

Study Setting

EHR Data

EHR data on 8,949 pregnant women from an urban academic medical center from 2015 to 2017 were extracted. The cohort inclusion and exclusion criteria are described in Figure 1. We excluded patients whose ages were below 18 or above 45, had no encounter recorded in the EHR from 1 year prior to pregnancy to 1 year after delivery, or missing home locations information. We extracted patient information including gender, age, race, ethnicity, body mass index (BMI), marital status, outpatient and inpatient diagnoses, outpatient and inpatient prescription medication orders, and corresponding encounter dates from the EHR data. Patient age was calculated as the time difference between the birth date and first prenatal checkup date. The gestational week was calculated using the date of delivery and the specific gestational age at prenatal checkup. Marital status was defined as single (single, divorced, widowed, unknown), and married, as extracted from unstructured clinical notes using regular expression. The trimester of each event was determined using the difference in time between each event and delivery. All diagnoses were represented as Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) codes.(93) Anatomical Therapeutic Chemical (ATC) Classification System was used to standardize the specific drug prescription and dosage information.(94) The primary outcome of PPD was defined as having at least one diagnosis of depression within 1 year after childbirth based on SNOMED codes [see Additional file 1].

Built Environment Data

Accessibility to public transportation

Three indicators were defined to measure the accessibility to public transportation and active transportation facilities: the number of bus stops within 500-meter radius, the number of subway stations within the 500-meter radius, and the length of bike paths within the 500-meter radius. The spatial data on public transportation and bike facilities were obtained in shapefile formats from New York State.(95) We used ArcGIS 10.6 spatial analysis tools to count the number of bus stops and subway stations within each 500-meter radius around each patients' home location and also to measure the length of bike paths within the 500-meter radius.

Exposure to Traffic

We obtained traffic data from the New York activity-based travel demand model referred to as "New York Best Practice Model (NYBPM)."(96) The model predicts daily traffic volume in each roadway link for the different types of vehicles by two categories: light- (passenger vehicles and taxis) and heavy-duty (buses and trucks) vehicles for their different levels of health impacts.(97) The vehicle kilometer traveled (VKT) within the 500-meter radius was then calculated based on the distance that vehicle pollution concentration reaches the background level.(98) VKT is calculated by multiplying traffic volume by the distance of travel, representing the amount of traffic activity.

Land Use

Five indicators were defined to measure the role of land use: entropy-based land use mix (LUM) index, retail floor area ratio (RetFAR), street connectivity, and sidewalk availability. The variables measure the availability and variety of destinations within 500 meters of the subject's home location. The land use data including information about land use class and parcel area at the parcel level were extracted from the parcel shapefile obtained from New York State.(95) The LUM index within 500-m radius measures the heterogeneity of land use, such as residential, commercial, retail, and industrial, within the radius.(99) The LUM index ranges between 0 to 1, where 0 represents homogeneity and 1 represents maximum heterogeneity.(99) Higher LUM values indicate higher walkability of the area. The RetFAR is the retail building floor area divided by the retail land area within the 250-m radius.(99) Examples with higher and lower RetFAR are multi-floor departmental stores and open-style outlets, respectively. The number of intersections within the 500-meter radius is another land use indicator used to measure the walkability of the neighborhood.(100) The number of intersections was extracted from the transportation network developed for the NYBPM travel demand model. To calculate the sidewalk area within the 500-meter radius, we used the sidewalk shapefiles as a measure of the accessibility of subjects to the walking facilities.(96)

Air pollution

Average daily particulate matter ($PM_{2.5}$) and ozone (O_3) concentrations at the census tract level for the period of 2015-2017 were obtained from the Center for Air, Climate and Energy Solutions which applied Land Use Regression (LUR) models to estimate every subject's exposure to air pollution.(101) $PM_{2.5}$ and O_3 together could represent both regional background and hotspot air pollution levels.

Other Social Determinants of Health (SDoH)

Lastly, SDoH information at the census-tract (11-digit Federal Information Processing Standard code) level were extracted using the FACETS dataset.(102) Variables used in the analysis included census-tract level average percent of college degree, GINI index, felony rate, and uninsured percentage from American Community Survey, a binary indicator of low access to healthy food within half mile from the Food Access Research Atlas, United States Department of Agriculture, the population-weighted distance to closest 7 parks from the Centers for Disease Control and Prevention, and lastly walk score scales the from Rundle-Columbia Built Environment and Health Research Group.

Patterns of Prenatal Care

We extracted the health and healthcare utilization information during the prenatal period for each patient from the EHR data. Patients who had similar overall prenatal care patterns were categorized into clusters as having experienced generally similar prenatal events. The similarity between pairs of patients were measured using the longest common subsequence (LCS) distance. LCS measures the longest overlap that 2 sequences have in common; thus, larger LCS indicates a more similar course of the clinical events. In this study, we compared the sequence of each patient's clinical events (e.g., encounters, diagnoses, prescription medications) to others in the cohort to generate pairs of LCS distances. Based on the similarity, the categorization of patients was performed using the hierarchical clustering algorithm, a well-established machine learning method for detecting underlying clusters in a population.(91) The final number and size of the clusters were determined using Silhouette value.(91) This method was previously used to mine EHR data to identify health and healthcare utilization patterns among patients with chronic kidney disease, heart failure, and undifferentiated abdominal pain.(91, 103, 104) An example of the sequences used for categorization is given in the Additional file 2.

Because of the large number (n>6,000) of unique clinical events recorded in the EHR data, we limited the pattern mining to focus on variables that were found to be most predictive of PPD in a related work preparatory to this study.(105) The list of variables, including complications during pregnancy and medication usage, are shown in Additional file 3. The cluster analysis was done in Python 3.6.5 and R 4.0.0.

Statistical Analysis

The distribution of study variables described in sections EHR Data and Built Environment Data (Table 1) were assessed within each identified cluster. Multivariate Imputation by Chained Equations (MICE) was used to address the missing value issue.(106) We further studied the relationship between prenatal care, as reflected by the cluster membership, the built environment characteristics, and incidence of PPD using structural equation models (SEMs).(92) Two SEMs were constructed for the primary and secondary outcomes separately. All independent variables were considered, but removed if there was multicollinearity as determined by variable inflation factor larger than 10. Statistical analysis was done using Stata/IC 16.0 and R 4.0.0. We applied Chi-square tests for categorical variables and analysis of variance (ANOVA) for continuous variables to compare the differences across clusters. P-value of 0.05 was used as the significance threshold.

Results

Table 1 shows the descriptive statistics of the study cohort where continuous variables are presented as mean (standard deviation (SD)), and categorical variables are presented as N (% in total cohort). The average age of our patient population was 33.7 years (SD=4.59). Nearly half (49.27%) of the patients were White, and majority were married (86.7%) and had Commercial insurances (84.1%). Over 3% of the cohort were diagnosed with PPD. A total of 3,922 (43.6%) and 482 (5.4%) patients had at least one ED visit pre- and post-delivery.

We identified 3 clusters with 1,955 (cluster 1), 4,188 (cluster 2), and 2,949 (cluster 3) patients, respectively, based on their clinical event sequences. For the primary outcome of PPD, 6.65% of the women in cluster 1 had a diagnosis of PPD within 1 year after childbirth, which was higher than clusters 2 (2.67%) and 3 (1.12%) (P<.05). Table 2 presents the distribution of demographics, medications, diagnoses, and built environment factors that were significantly different across the three clusters. The mean (SD) age across three clusters were 35.01 (4.73) years, 33.78 (4.29) years and 32.68 (4.66) years, respectively (P<.001). There were more unmarried patients in cluster 1 than the other two clusters (P<.001). In addition, the number of ED visits in both the pre- and post-delivery periods in the cluster 1 were significantly higher (P<.05) than the other clusters. In terms of medication prescriptions, we observed significantly higher rates of prescription medications in cluster 1, such as analgesics, antipyretics and opioids (P<.001). Further, more patients in cluster 1 had complications during pregnancy, unplanned pregnancies, high-risk pregnancy, abnormal glucose level, elderly primigravida and advanced maternal age gravidas than the other two clusters (P<.001). Additional file 2 showcases sequential patterns in the prenatal care identified from the

study data.

Table 3 displays the results from the SEM for the outcome of PPD. Regarding the primary outcome, patients in clusters 1 (odds ratio=6.3, P<.001) and 2 (odds ratio=2.43, P<.001) are more likely to have a diagnosis PPD within 12 months after childbirth than women in cluster 3. Relative to cluster 3, patients in cluster 1 are more likely to have patients living in census tract that have lower PM 2.5 (odds ratio=0.858, P=.02), lower retail floor area ratio (odds ratio=0.882, P=.03), lower LUM (odds ratio=0.508, P<.001), higher GINI (odds ratio=4.317, P=0.002), and higher college degree percentage (odds ratio=4.401, P<.001). Patients are also more likely to be older in age (odds ratio=1.115, P <.001) and not married (odds ratio=0.404, P<.001). Relative to cluster 3, patients in cluster 2 are more likely to have patients living in census tract that have lower PM 2.5 (odds ratio=0.890, P=0.03), lower retail floor area ratio (odds ratio=0.867, P=.001), lower GINI (odds ratio=0.412, P=0.02), and higher college degree percentage (odds ratio=4.996, P<.001). Patients are also moderately more likely to be older in age (odds ratio=1.046, P<.001) and not married (odds ratio=0.560, P<.001). Race and insurance types (commercial, Medicaid, Other including Medicare) were not significantly associated with the cluster membership in the models although unadjusted association was significant.

Within each cluster, we further examined the characteristics of PPD cases as shown in Additional file 4. The association between PPD and the built environment factors were examined and shown in Additional file 5. The factors that were significantly associated with increased risk for PPD were the number of intersections within 500-m radius, the number of bus stops within 500-m radius, and retail floor area ratio, while adjusting for felony rates and GINI index which were also significant in the model.

Discussion

There were two major findings in this study. Three clusters of prenatal health and healthcare utilization patterns were discovered from a cohort of women whose pregnancies were managed entirely or partially in an urban academic medical center in 2015 to 2017. The distribution of the primary and secondary outcomes of PPD were significantly different across the clusters. Clinically, the clusters differed in maternal age, BMI, marital status, medication use, chronic conditions, and complications during pregnancy. In addition, we found that the cluster

membership was associated with built environment factors related to walkability, access to retail resources, air quality, as well as neighborhood felony rates, and neighborhood income equality. These findings contribute to the growing body of evidence that the built environment in the community confers an impact on the trajectories of health and health service utilization during pregnancy.

The associations found between retail, land-use and the study outcomes among the pregnant cohort are novel and important contributions to the literature. Retail floor area ratio is indicative of pedestrian-orientated design and higher walkability. The mixed land use and more retail access may be a proxy for the connectedness of the neighborhood in providing community support to women. These community resources potentially lead to increased opportunities for social contact, lower stress levels, and higher physical activity levels, which is consistent with previous literature tying maternal mental health to green space.(67, 68) Air quality has been linked with adverse birth outcomes including preterm birth and miscarriages in previous literature.(67) However, we found that lower PM 2.5 concentration to be associated with clusters with higher PPD incidences in contrary to previous literature. In our urban study setting, PM 2.5 concentration is highest in the most affluent area and becomes lower as we move out to other parts of the study setting. Therefore, our findings on the association of poor air quality with higher incidence PPD case potentially reflect patient cohorts who are predominantly in or outside the most affluent part of the city who have better access to mental health reporting and care. Patterns learned from this study may inform expecting and new mothers, their care providers, as well as guideline and policy makers, to better prepare and navigate pregnancy and postpartum care. Additionally, our findings may have implications for policies during the current COVID-19 pandemic as our communities and their stores face significant changes.

There are limitations in the study. All diagnoses in the study were defined using diagnostic codes. Therefore, missed and under-diagnosis of health conditions during pregnancy, including PPD, is a crucial limitation. It is possible that this study missed PPD patients who did not disclose symptoms due to stigma against mental health, and patients who were diagnosed outside of our health system. The under- and mis-diagnosis may be more prevalent among women who live in low-income neighborhoods. Some of these limitations may be addressed in future work by patient interviews and questionnaires. Additionally, the application of natural language processing on unstructured clinical notes may allow us to elicit underdiagnosed and missed PPD as well as other conditions. Moreover, we were

not able to address the possible reporting bias in our study population with respect to information such as race and marital status. Nearly 15% of the racial information was unknown from the EHR data. Future studies may explore the leveraging of patient-reported outcome data in overcoming this limitation. Furthermore, in analyzing the medication data, we did not consider the dose-response relationship between medications and the outcome as prescription fill information was not available. Detailed medication dose and frequency information can be analyzed in future work if pharmacy claims data become available. Lastly, while this study used data from a single health system in NYC, further work will aim to validate our findings using EHR data from other institutions and across different cities in the US.

Conclusion

We found that poor-quality built environment is associated with variability in prenatal care and maternal mental health outcomes in a large retrospective cohort study using EHR data. Findings from this study may inform healthcare providers and public health policymakers in understanding modifiable risk factors that are associated with poor pregnancy care and outcomes.

Reference

1. Stewart DE, Vigod S. Postpartum Depression. N Engl J Med. 2016;375(22):2177-86.
2. Hahn-Holbrook J, Cornwell-Hinrichs T, Anaya I. Economic and Health Predictors of National Postpartum Depression Prevalence: A Systematic Review, Meta-analysis, and Meta-Regression of 291 Studies from 56 Countries. Front Psychiatry. 2017;8:248.
3. Wisner KL, Sit DK, McShea MC, Rizzo DM, Zoretich RA, Hughes CL, et al. Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. JAMA Psychiatry. 2013;70(5):490-8.
4. O'Connor E, Senger CA, Henninger ML, Coppola E, Gaynes BN. Interventions to Prevent Perinatal Depression: Evidence Report and Systematic Review for the US Preventive Services Task Force. Jama. 2019;321(6):588-601.
5. Biaggi A, Conroy S, Pawlby S, Pariante CM. Identifying the women at risk of antenatal anxiety and depression: A systematic review. J Affect Disord. 2016;191:62-77.
6. Jacques N, de Mola CL, Joseph G, Mesenburg MA, da Silveira MF. Prenatal and postnatal maternal depression and infant hospitalization and mortality in the first year of life: A systematic review and meta-analysis. J Affect Disord. 2019;243:201-8.
7. Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, et al. Effects of perinatal mental disorders on the fetus and child. Lancet. 2014;384(9956):1800-19.
8. Weobong B, ten Asbroek AH, Soremekun S, Gram L, Amenga-Etego S, Danso S, et al. Association between probable postnatal depression and increased infant mortality and morbidity:

findings from the DON population-based cohort study in rural Ghana. BMJ Open. 2015;5(8):e006509.
9. Field T. Postpartum depression effects on early interactions, parenting, and safety practices: a review. Infant Behav Dev. 2010;33(1):1-6.
10. Meltzer-Brody S, Howard LM, Bergink V, Vigod S, Jones I, Munk-Olsen T, et al. Postpartum psychiatric disorders. Nat Rev Dis Primers. 2018;4:18022.
11. Serati M, Redaelli M, Buoli M, Altamura AC. Perinatal Major Depression Biomarkers: A systematic review. J Affect Disord. 2016;193:391-404.
12. Committee A. ACOG Committee Opinion No. 757: Screening for Perinatal Depression. Obstet Gynecol. 2018;132(5):e208-e12.
13. Earls MF, Yogman MW, Mattson G, Rafferty J. Incorporating Recognition and Management of Perinatal Depression Into Pediatric Practice. Pediatrics. 2019;143(1).
14. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry. 1987;150:782-6.
15. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care. 2004;42(12):1194-201.
16. Cohen LS, Altshuler LL, Harlow BL, Nonacs R, Newport DJ, Viguera AC, et al. Relapse of major depression during pregnancy in women who maintain or discontinue antidepressant treatment. Jama. 2006;295(5):499-507.
17. Gjerdingen DK, Yawn BP. Postpartum depression screening: importance, methods, barriers, and recommendations for practice. J Am Board Fam Med. 2007;20(3):280-8.
18. Beck CT, Gable RK. Postpartum Depression Screening Scale: development and psychometric testing. Nurs Res. 2000;49(5):272-82.
19. Wang S, Pathak J, Zhang Y. Using Electronic Health Records and Machine Learning to Predict Postpartum Depression. Stud Health Technol Inform. 2019;264:888-92.
20. Loudon H, Nentin F, Silverman ME. Using clinical decision support as a means of implementing a universal postpartum depression screening program. Arch Womens Ment Health. 2016;19(3):501-5.
21. Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. Jama. 2018;320(21):2199-200.
22. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc. 2017;24(1):198-208.
23. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature. 2019;572(7767):116-9.
24. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1:18.
25. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25(3):433-8.
26. Bishop CM. Pattern recognition and machine learning: Springer-Verlag New York; 2006.
27. Tortajada S, García-Gomez JM, Vicente J, Sanjuán J, de Frutos R, Martín-Santos R, et al. Prediction of postpartum depression using multilayer perceptrons and pruning. Methods Inf Med. 2009;48(3):291-8.

28. Jiménez-Serrano S, Tortajada S, García-Gómez JM. A Mobile Health Application to Predict Postpartum Depression Based on Machine Learning. Telemed J E Health. 2015;21(7):567-74.
29. Zhang W, Liu H, Silenzio VMB, Qiu P, Gong W. Machine Learning Models for the Prediction of Postpartum Depression: Application and Comparison Based on a Cohort Study. JMIR Med Inform. 2020;8(4):e15516.
30. 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):1515-20.
31. Petersen I, Peltola T, Kaski S, Walters KR, Hardoon S. Depression, depressive symptoms and treatments in women who have recently given birth: UK cohort study. BMJ Open. 2018;8(10):e022152.
32. Schofield P, Das-Munshi J, Mathur R, Congdon P, Hull S. Does depression diagnosis and antidepressant prescribing vary by location? Analysis of ethnic density associations using a large primary-care dataset. Psychol Med. 2016;46(6):1321-9.
33. Kaushal R, Hripcsak G, Ascheim DD, Bloom T, Campion TR, Jr., Caplan AL, et al. Changing the research landscape: the New York City Clinical Data Research Network. J Am Med Inform Assoc. 2014;21(4):587-90.
34. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54-60.
35. Canada S. Mental disorders, conditions, and ICD-9 and ICD-10-CA codes: Statistics Canada; 2015 [Available from: https://www150.statcan.gc.ca/n1/pub/82-622-x/2011006/tbl/tbla-eng.htm.
36. Rodríguez P, Bautista MA, Gonzàlez J, Escalera S. Beyond one-hot encoding: Lower dimensional target embedding. Image and Vision Computing. 2018;75:21-31.
37. T. L, G. T, editors. SFS feature selection technique for multistage emotion recognition. 2015 IEEE 3rd Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE); 2015 13-14 Nov. 2015.
38. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997;30(7):1145-59.
39. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
40. Rufibach K. Use of Brier score to assess binary predictions. J Clin Epidemiol. 2010;63(8):938-9; author reply 9.
41. Stewart CC, Lu CY, Yoon TK, Coleman KJ, Crawford PM, Lakoma MD, et al. Impact of ICD-10-CM Transition on Mental Health Diagnoses Recording. EGEMS (Wash DC). 2019;7(1):14.
42. Forman DN, Videbech P, Hedegaard M, Salvig JD, Secher NJ. Postpartum depression: identification of women at risk. Brit J Obstet Gynaec. 2000;107(10):1210-7.
43. Le Donne M, Mento C, Settineri S, Antonelli A, Benvenga S. Postpartum Mood Disorders and Thyroid Autoimmunity. Front Endocrinol (Lausanne). 2017;8:91.
44. Strapasson MR, Ferreira CF, Ramos JGL. Associations between postpartum depression and hypertensive disorders of pregnancy. Int J Gynecol Obstet. 2018;143(3):367-73.
45. Barsky AJ, Cleary PD, Coeytaux RR, Ruskin JN. Psychiatric disorders in medical outpatients complaining of palpitations. J Gen Intern Med. 1994;9(6):306-13.

46. Alijaniha F, Noorbala A, Afsharypuor S, Naseri M, Fallahi F, Mosaddegh M, et al. Relationship Between Palpitation and Mental Health. Iran Red Crescent Me. 2016;18(3).
47. Yudofsky SC. Beta-blockers and depression. The clinician's dilemma. JAMA : the journal of the American Medical Association. 1992;267(13):1826-7.
48. Gerstman BB, Jolson HM, Bauer M, Cho P, Livingston JM, Platt R. The incidence of depression in new users of beta-blockers and selected antihypertensives. J Clin Epidemiol. 1996;49(7):809-15.
49. Ozdemir PG, Karadag AS, Selvi Y, Boysan M, Bilgili SG, Aydin A, et al. Assessment of the effects of antihistamine drugs on mood, sleep quality, sleepiness, and dream anxiety. Int J Psychiat Clin. 2014;18(3):161-8.
50. Xu H, Ding Y, Ma Y, Xin XL, Zhang DF. Cesarean section and risk of postpartum depression: A meta-analysis. J Psychosom Res. 2017;97:118-26.
51. Carter FA, Frampton CMA, Mulder RT. Cesarean section and postpartum depression: A review of the evidence examining the link. Psychosom Med. 2006;68(2):321-30.
52. Sheen JJ, Smith HA, Tu B, Liu Y, Sutton D, Bernstein PS. Risk Factors for Postpartum Emergency Department Visits in an Urban Population. Matern Child Hlth J. 2019;23(4):557-66.
53. Long MM, Cramer RJ, Jenkins J, Bennington L, Paulson JF. A systematic review of interventions for healthcare professionals to improve screening and referral for perinatal mood and anxiety disorders. Arch Women Ment Hlth. 2019;22(1):25-36.
54. Goodman JH, Tyer-Viola L. Detection, Treatment, and Referral of Perinatal Depression and Anxiety by Obstetrical Providers. J Womens Health. 2010;19(3):477-90.
55. Camdeviren HA, Yazici AC, Akkus Z, Bugdayci R, Sungur MA. Comparison of logistic regression model and classification tree: An application to postpartum depression data. Expert Systems with Applications. 2007;32(4):987-94.
56. S. N, A. P, N. R, A. B, K. S, K. C, editors. Boosting for Postpartum Depression Prediction. 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); 2017 17-19 July 2017.
57. Lee WC, Veeranki SP, Serag H, Eschbach K, Smith KD. Improving the Collection of Race, Ethnicity, and Language Data to Reduce Healthcare Disparities: A Case Study from an Academic Medical Center. Perspect Health Inf Manag. 2016;13(Fall):1g.
58. Sholle ET, Pinheiro LC, Adekkanattu P, Davila MA, Johnson SB, Pathak J, et al. Underserved populations with missing race ethnicity data differ significantly from those with structured race/ethnicity documentation. Journal of the American Medical Informatics Association : JAMIA. 2019;26(8-9):722-9.
59. Koh HK. A 2020 Vision for Healthy People. New England Journal of Medicine. 2010;362(18):1653-6.
60. Chaiyachati KH, Hom JK, Hubbard RA, Wong C, Grande D. Evaluating the association between the built environment and primary care access for new Medicaid enrollees in an urban environment using Walk and Transit Scores. Prev Med Rep. 2018;9:24-8.
61. Beutel ME, Brahler E, Ernst M, Klein E, Reiner I, Wiltink J, et al. Noise annoyance predicts symptoms of depression, anxiety and sleep disturbance 5 years later. Findings from the Gutenberg Health Study. Eur J Public Health. 2020;30(3):516-21.
62. Galea S, Ahern J, Rudenstine S, Wallace Z, Vlahov D. Urban built environment and depression: a multilevel analysis. J Epidemiol Community Health. 2005;59(10):822-7.
63. Emeruwa UN, Ona S, Shaman JL, Turitz A, Wright JD, Gyamfi-Bannerman C, et al. Associations Between Built Environment, Neighborhood Socioeconomic Status, and SARS-CoV-

2 Infection Among Pregnant Women in New York City. JAMA : the journal of the American Medical Association. 2020.
64. Mullings JA, McCaw-Binns AM, Archer C, Wilks R. Gender differences in the effects of urban neighborhood on depressive symptoms in Jamaica. Rev Panam Salud Publ. 2013;34(6):385-92.
65. Veldhuis CB, Maki P, Molina K. Psychological and neighborhood factors associated with urban women's preventive care use. J Behav Med. 2020;43(3):346-64.
66. Guglielminotti J, Landau R, Wong CA, Li G. Patient-, Hospital-, and Neighborhood-Level Factors Associated with Severe Maternal Morbidity During Childbirth: A Cross-Sectional Study in New York State 2013-2014. Matern Child Health J. 2019;23(1):82-91.
67. McEachan RRC, Prady SL, Smith G, Fairley L, Cabieses B, Gidlow C, et al. The association between green space and depressive symptoms in pregnant women: moderating roles of socioeconomic status and physical activity. J Epidemiol Commun H. 2016;70(3):253-9.
68. Nichani V, Dirks K, Burns B, Bird A, Grant C. Green Space and Depression during Pregnancy: Results from the Growing Up in New Zealand Study. Int J Environ Res Public Health. 2017;14(9).
69. Glance LG, Dick AW, Glantz JC, Wissler RN, Qian F, Marroquin BM, et al. Rates Of Major Obstetrical Complications Vary Almost Fivefold Among US Hospitals. Health affairs. 2014;33(8):1330-6.
70. Grobman WA, Bailit JL, Rice MM, Wapner RJ, Varner MW, Thorp JM, et al. Can differences in obstetric outcomes be explained by differences in the care provided? The MFMU Network APEX study. American Journal of Obstetrics and Gynecology. 2014;211(2).
71. Farr SL, Dietz PM, Rizzo JH, Vesco KK, Callaghan WM, Bruce FC, et al. Health Care Utilisation in the First Year of Life Among Infants of Mothers With Perinatal Depression or Anxiety. Paediatr Perinat Ep. 2013;27(1):81-8.
72. Cunningham SD, Magriples U, Thomas JL, Kozhimannil KB, Herrera C, Barrette E, et al. Association Between Maternal Comorbidities and Emergency Department Use Among a National Sample of Commercially Insured Pregnant Women. Academic Emergency Medicine. 2017;24(8):940-7.
73. Kilfoyle KA, Vrees R, Raker CA, Matteson KA. Nonurgent and urgent emergency department use during pregnancy: an observational study. American journal of obstetrics and gynecology. 2017;216(2):181. e1-. e7.
74. D'Ascoli PT, Alexander GR, Petersen DJ, Kogan MD. Parental factors influencing patterns of prenatal care utilization. J Perinatol. 1997;17(4):283-7.
75. Akaraci S, Feng XQ, Suesse T, Jalaludin B, Astell-Burt T. A Systematic Review and Meta-Analysis of Associations between Green and Blue Spaces and Birth Outcomes. Int J Env Res Pub He. 2020;17(8).
76. Giurgescu C, Zenk SN, Templin TN, Engeland CG, Dancy BL, Park CG, et al. The Impact of Neighborhood Environment, Social Support, and Avoidance Coping on Depressive Symptoms of Pregnant African-American Women. Womens Health Issues. 2015;25(3):294-302.
77. Jacques N, de Mola CL, Josephc G, Mesenburg MA, da Silveira MF. Prenatal and postnatal maternal depression and infant hospitalization and mortality in the first year of life: A systematic review and meta-analysis. J Affect Disorders. 2019;243:201-8.
78. Weobong B, ten Asbroek AHA, Soremekun S, Gram L, Amenga-Etego S, Danso S, et al. Association between probable postnatal depression and increased infant mortality and morbidity: findings from the DON population-based cohort study in rural Ghana. Bmj Open. 2015;5(8).

79. Field T. Postpartum depression effects on early interactions, parenting, and safety practices: A review. Infant Behav Dev. 2010;33(1):1-6.
80. Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, et al. Effects of perinatal mental disorders on the fetus and child. Lancet. 2014;384(9956):1800-19.
81. Moore Simas TA, Huang MY, Packnett ER, Zimmerman NM, Moynihan M, Eldar-Lissai A. Matched cohort study of healthcare resource utilization and costs in young children of mothers with postpartum depression in the United States. Journal of medical economics. 2020;23(2):174-83.
82. Silverman ME, Reichenberg A, Savitz DA, Cnattingius S, Lichtenstein P, Hultman CM, et al. The risk factors for postpartum depression: A population-based study. Depress Anxiety. 2017;34(2):178-87.
83. Howard LM, Molyneaux E, Dennis CL, Rochat T, Stein A, Milgrom J. Non-psychotic mental disorders in the perinatal period. Lancet. 2014;384(9956):1775-88.
84. Chen HL, Cai JY, Zha ML, Shen WQ. Prenatal smoking and postpartum depression: a meta-analysis. J Psychosom Obstet Gynaecol. 2019;40(2):97-105.
85. OHara MW, Swain AM. Rates and risk of postpartum depression - A meta-analysis. Int Rev Psychiatr. 1996;8(1):37-54.
86. Zhang SM, Wang LS, Yang TB, Chen LZ, Qiu X, Wang TT, et al. Maternal violence experiences and risk of postpartum depression: A meta-analysis of cohort studies. Eur Psychiat. 2019;55:90-101.
87. Norhayati MN, Hazlina NHN, Asrenee AR, Emilin WMAW. Magnitude and risk factors for postpartum symptoms: A literature review. J Affect Disorders. 2015;175:34-52.
88. Feng XQ, Astell-Burt T. Residential green space quantity and quality and symptoms of psychological distress: a 15-year longitudinal study of 3897 women in postpartum. Bmc Psychiatry. 2018;18.
89. He SY, Smargiassi A, Low N, Bilodeau-Bertrand M, Ayoub A, Auger N. Residential noise exposure and the longitudinal risk of hospitalization for depression after pregnancy: Postpartum and beyond. Environ Res. 2019;170:26-32.
90. Schinasi LH, Auchincloss AH, Forrest CB, Roux AVD. Using electronic health record data for environmental and place based population health research: a systematic review. Ann Epidemiol. 2018;28(7):493-502.
91. Zhang Y, Padman R, Patel N. Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data. Journal of biomedical informatics. 2015.
92. Huber C. Introduction to Structural Equation Modeling Using Stata. California Association for Institutional Research. 2014.
93. Odigie E, Lacson R, Raja A, Osterbur D, Ip I, Schneider L, et al. Fast Healthcare Interoperability Resources, Clinical Quality Language, and Systematized Nomenclature of Medicine-Clinical Terms in Representing Clinical Evidence Logic Statements for the Use of Imaging Procedures: Descriptive Study. JMIR Med Inform. 2019;7(2):e13590.
94. Ronning M, Blix HS, Harbo BT, Strom H. Different versions of the anatomical therapeutic chemical classification system and the defined daily dose - are drug utilisation data comparable? Eur J Clin Pharmacol. 2000;56(9-10):723-7.
95. State NY. The Official Website of New York State [Available from: https://www.ny.gov/.
96. Vovsha P, Petersen E, Donnelly R. Microsimulation in travel demand modeling: Lessons learned from the New York best practice model. Transportation Research Record: Journal of the Transportation Research Board. 2002(1805):68-77.

97. Pollution HEIPotHEoT-RA. Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects: Health Effects Institute; 2010.
98. Karner AA, Eisinger DS, Niemeier DA. Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. Environ Sci Technol. 2010;44(14):5334-44.
99. Frank LD, Sallis JF, Conway TL, Chapman JE, Saelens BE, Bachman W. Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality. Journal of the American planning Association. 2006;72(1):75-87.
100. Sugiyama T, Leslie E, Giles-Corti B, Owen N. Associations of neighbourhood greenness with physical and mental health: do walking, social coherence and local social interaction explain the relationships? J Epidemiol Commun H. 2008;62(5).
101. Kim S-Y, Bechle M, Hankey S, Sheppard L, Szpiro A, Marshall J, editors. A Parsimonious Approach to National Prediction: Criteria Pollutants in the Contiguous US, 1979-2015. ISEE Conference Abstracts; 2018.
102. Cantor MN, Chandras R, Pulgarin C. FACETS: using open data to measure community social determinants of health. Journal of the American Medical Informatics Association. 2017;0(0).
103. Zhang Y, Padman R, Epner P, Bauer V, Solomonides A, Rao G. Identifying Diagnostic Paths for Undifferentiated Abdominal Pain from Electronic Health Record Data. AMIA Jt Summits Transl Sci Proc. 2018;2017:290-9.
104. Movahedi F, Kormos RL, Lohmueller L, Seese L, Kanwar M, Murali S, et al. Sequential pattern mining of longitudinal adverse events after Left Ventricular Assist Device implant. IEEE journal of biomedical and health informatics. 2019.
105. Wang S, Pathak J, Zhang Y. Using Electronic Health Records and Machine Learning to Predict Postpartum Depression. Studies in health technology and informatics. 2019;264:888-92.
106. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Meth Psych Res. 2011;20(1):40-9.

Figure 1. Study cohort inclusion and exclusion criteria

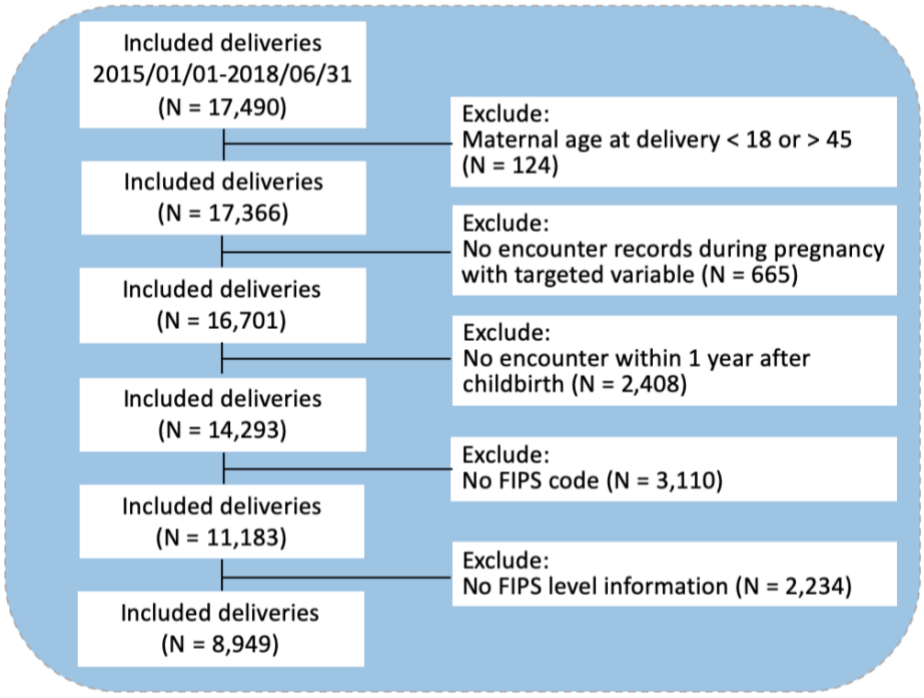


Table 1. Descriptive statistics of the study cohort

Variables	**Values**
Demographics	
Age, mean (SD), year	33.69 (4.59)
Pre-pregnancy BMI, mean (SD), kg/m^2	23.77 (4.31)
Gestational Week, mean (SD), week	38.69 (2.09)
Race, No. (%)	
White	4409 (49.27)
Asian	1689 (18.87)
Black or African American	560 (6.26)
Other	976 (10.91)
Unknown	1315 (14.69)
Marital Status, No. (%)	
Single	1193 (13.33)
Married	7756 (86.67)
Cesarean Section, No. (%)	
Yes	1878 (20.99)

No	7071 (79.01)
Insurance, No. (%)	
Commercial	7519 (84.02)
Medicaid	1226 (13.70)
Other	204 (2.28)
Built Environment	
Number of bus stops within 500 m radius, mean (SD)	25.26 (10.0)
Number of subway stations within 500 m radius, mean (SD)	1.81 (1.83)
Parks Area within 500 m radius, mean (SD), m^2	463112.43 (660506.3)
Bike Path Length within 500 m radius, mean (SD), m	29070.94 (15172.89)
VKT of light vehicles within 500 m radius, mean (SD), 100,000 units	3283.87 (2242.98)
VKT of heavy vehicles within 500 m radius, mean (SD), 10,000 units	3608.43 (2516.02)
LUM index within 500 m radius, mean (SD)	0.64 (0.17)
RetFar within 500 m radius, mean (SD)	0.24 (0.23)
Number of Intersections within 500 m radius, mean (SD)	12.06 (7.76)
Sidewalk Area within 500 m radius, mean (SD), 1000 m^2	907.77 (208.53)
Ozone Concentration, mean (SD), $\mu g/m^3$	46.56 (0.50)
$PM_{2.5}$ Concentration, mean (SD), $\mu g/m^3$	9.28 (0.47)
Percent of Colleges Degree, mean (SD), %	35.79 (11.49)
Average Poverty Rate, mean (SD), %	1.62 (2.15)
Average Respiratory Hazard Index, mean (SD)	4.51 (1.16)
Low Access to Healthy Food, No. (%)	297 (3.32)
Uninsured Percentage, mean (SD), %	8.26 (5.60)
Postpartum Depression	
Yes, No. (%)	273 (3.05)
Average number of ED visits per patient	
Pre-delivery (N=3900, 43.58%), mean (SD)	0.74 (1.16)
Post-delivery (N=482, 5.39%), mean (SD)	0.07 (0.31)

Table 2. Associations between cluster membership and clinical variables used for clustering

Variables	**Cluster**			**P-value**
	1 (N=1934)	**2 (N=4129)**	**3 (N=2886)**	

Demographics				
Age, mean (SD), year	35.01 (4.73)	33.78 (4.29)	32.68 (4.66)	<.001
Pre-pregnancy BMI, mean (SD), kg/m^2	24.24 (5.19)	23.55 (4.32)	23.77 (3.54)	<.001
Gestational Week, mean (SD), week	38.58 (2.12)	38.83 (1.92)	38.55 (2.26)	<.001
Race, no. (%)				
White	1078 (55.74)	2149 (52.05)	1182 (40.96)	<.001
Asian	280 (14.48)	679 (16.44)	730 (25.29)	
Black or African American	145 (7.50)	260 (6.30)	155 (5.37)	
Other	229 (11.84)	477 (11.55)	270 (9.36)	
Unknown	202 (10.44)	564 (13.66)	549 (19.02)	
Marital Status, no. (%)				
Single	348 (17.99)	578 (14.0)	267 (9.25)	<.001
Married	1586 (82.01)	3551 (86.0)	2619 (90.75)	
Average Poverty Rate, mean (SD), %	1.35 (1.83)	1.42 (1.87)	2.07 (2.61)	<.001
Cesarean Section, no. (%)				
Yes	510 (26.37)	833 (20.17)	535 (18.54)	<.001
No	1424 (73.63)	3296 (79.83)	2351 (81.46)	
Insurance, no. (%)				
Commercial	1603 (82.89)	3492 (84.57)	2424 (83.99)	.45
Medicaid	283 (14.63)	552 (13.37)	391 (13.55)	
Other (Medicare, Self-pay, Unknown)	48 (2.48)	85 (2.06)	71 (2.46)	
ED Visits per patient				
Pre-delivery (within 1-year), mean (SD)	1.12 (1.54)	0.68 (1.01)	0.56 (0.97)	<.001
Post-delivery (within 6-months), mean (SD)	0.10 (0.37)	0.06 (0.29)	0.05 (0.28)	<.001
Medication Prescriptions				
Other Analgesics and Antipyretics, no. (%)	324 (16.75)	534 (12.93)	324 (11.23)	<.001
Opioids, no. (%)	285 (14.74)	323 (7.82)	243 (8.42)	<.001
Thyroid Preparations, no. (%)	291 (15.05)	273 (6.61)	84 (2.91)	<.001

Drugs for Functional Gastrointestinal Disorders, no. (%)	171 (8.84)	235 (5.69)	150 (5.2)	<.001
Antiemetics and Antinauseants, no. (%)	170 (8.79)	242 (5.86)	145 (5.02)	<.001
Other Plain Vitamin Preparations, no. (%)	172 (8.89)	252 (6.10)	83 (2.88)	<.001
Antihistamines for Systemic Use, no. (%)	185 (9.57)	234 (5.67)	83 (2.88)	<.001
Beta-lactam Antibacterials, Penicillins, no. (%)	175 (9.05)	245 (5.93)	81 (2.81)	<.001
Progestogens, no. (%)	284 (14.68)	156 (3.78)	42 (1.46)	<.001
Direct Acting Antivirals, no. (%)	143 (7.39)	187 (4.53)	70 (2.43)	<.001
Diagnoses				
Normal Delivery, no. (%)	1435 (74.2)	3346 (81.04)	2310 (80.04)	<.001
Primigravida, no. (%)	1206 (62.36)	2453 (59.41)	1024 (35.48)	<.001
Complication Occurring During Pregnancy, no. (%)	887 (45.86)	1439 (34.85)	605 (20.96)	<.001
Unplanned Pregnancy, no. (%)	641 (33.14)	1178 (28.53)	742 (25.71)	<.001
Post-term Pregnancy, no. (%)	465 (24.04)	1116 (27.03)	532 (18.43)	<.001
Elderly Primigravida, no. (%)	674 (34.85)	935 (22.64)	360 (12.47)	<.001
High Risk Pregnancy, no. (%)	536 (27.71)	662 (16.03)	297 (10.29)	<.001
Abnormal Glucose Level, no. (%)	479 (24.77)	757 (18.33)	163 (5.65)	<.001
Advanced Maternal Age Gravida, no. (%)	416 (21.51)	675 (16.35)	222 (7.69)	<.001
Disorder of Pregnancy, no. (%)	342 (17.68)	499 (12.09)	276 (9.56)	<.001
Postpartum Depression				
Yes, no. (%)	130 (6.72)	110 (2.66)	33 (1.14)	<.001
No, no. (%)	1804 (93.28)	4019 (97.34)	2853 (98.86)	

Table 3. Built environment factors that are associated with cluster membership while controlling for social-demographic factors. OR: odds ratio

	Variable	OR	P-value
PPD	Cluster 1	6.3	<.001

	Cluster 2	2.43	<.001
Cluster 1 (vs. cluster 3)	Retail	0.882	.03
	PM2.5	0.858	.02
	Age	1.115	<.001
	Married	0.404	<.001
	LUM	0.508	<.001
	GINI	4.317	.002
	College	4.401	<.001
	_cons	0.069	<.001
Cluster 2 (vs. cluster 3)	Retail	0.867	.001
	PM2.5	0.890	.03
	Age	1.046	<.001
	Married	0.560	<.001
	LUM	0.749	.06
	GINI	0.412	.02
	College	4.996	<.001
	_cons	1.734	.33

Additional file 1 Definition of PPD based on SNOMED codes

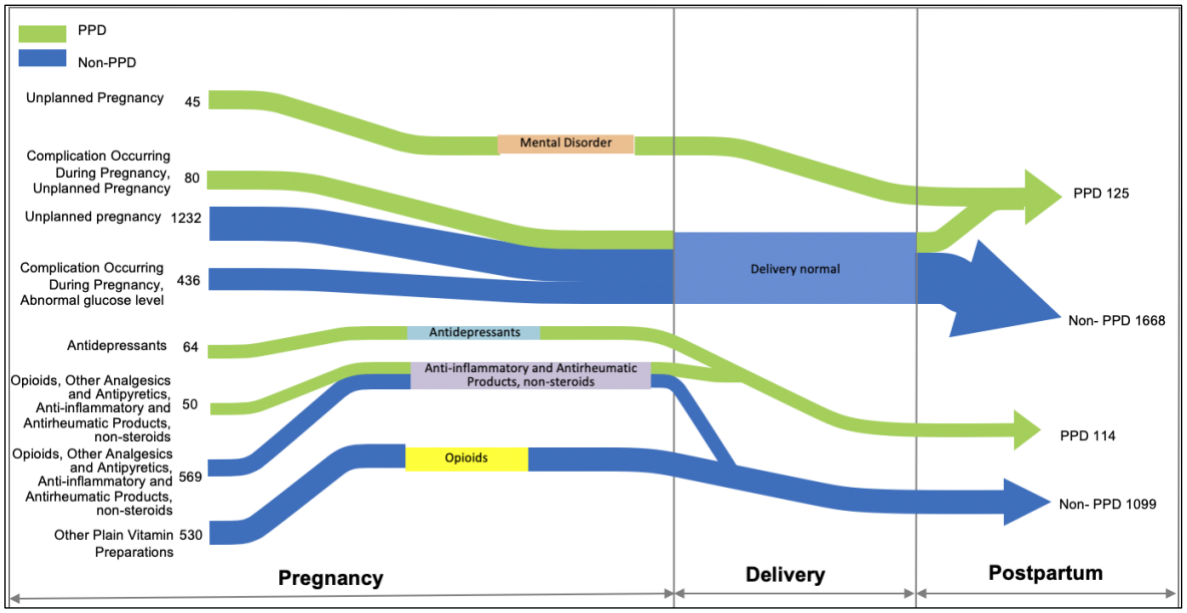
Concept name	**SNOMED code**
Acute depression	712823008
Adjustment disorder with depressed mood	57194009
Adjustment disorder with depressed mood in remission	698696007
Adjustment disorder with mixed anxiety and depressed mood	782501005
Anxiety	48694002/ 38237000
Anxiety disorder	197480006/ 191703000/ 65673007
Anxiety disorder in mother complicating childbirth	10743001000119100
Anxiety in pregnancy	94641000119109
Anxiety state	198288003/154882009
Anxiety state NOS	191711005
Anxiety states	268752000
Chronic anxiety	191708009
Depressed mood	366979004/41006004/367204005
Depressed mood with postpartum onset	704678007
Depression - postnatal	154889000
Depression NOS	307537002/154963001
Depressive conduct disorder	192605002
Depressive disorder in mother complicating pregnancy	94631000119100
Depressive disorder in remission	698957003
Generalized anxiety disorder	21897009/191706008/192401002
Major depression in full remission	63412003
Major depression in partial remission	30605009
Major depression in remission	42810003

Major depression single episode, in partial remission	70747007
Major depression, single episode	36923009
Major depression, melancholic type	320751009/62951006
Major depressive disorder	370143000
Major depressive disorder, single episode with atypical features	42925002
Major depressive disorder, single episode with melancholic features	63778009
Major depressive disorder, single episode with postpartum onset	25922000
Mild anxiety	70997004
Mild depression	310495003/154965008/390717003
Mild postnatal depression	237349002
Mild major depression	87512008
Minimal depression	718636001
Minimal major depression	720455008
Minimal major depression single episode	720454007
Minimal recurrent major depression	720451004
Minor depressive disorder	48589009
Mixed anxiety and depressive disorder	231504006/154964007/191707004/231504006/154964007/191707004
Postpartum depression	58703003
Perinatal depression	10211000132109
Postnatal depression	191740008
Postnatal depressive disorder	147016002
Mild postnatal depression	237349002
Moderate anxiety	61387006
Moderate depression	310496002/154919005/154966009
Moderate major depression	832007
Moderately severe depression	719593009
Moderately severe major depression	719592004
Moderately severe major depression single episode	720453001
Moderately severe recurrent major depression	720452006
Mood disorder with major depressive-like episode due to general medical condition	77486005
Severe major depression	450714000
Severe depression	310497006/154967000
Severe postnatal depression	237350002
Severe recurrent major depression	281000119103
Severe major depression, single episode	251000119105
Severe major depression without psychotic features	75084000
Severe recurrent major depression without psychotic features	36474008
Severe anxiety	80583007
Single episode of major depression in full remission	19527009
Single major depressive episode	268620009/192366006/
Single major depressive episode, in full remission	191606003
Single major depressive episode, in partial or unspecified remission	191605004
Single major depressive episode, mild	191601008
Single major depressive episode, moderate	191602001
Single major depressive episode, unspecified	191600009
Recurrent anxiety	191709001
Recurrent major depression	66344007
Recurrent major depression in full remission	46244001
Recurrent major depression in remission	68019004
Recurrent major depressive disorder with postpartum onset	71336009
Recurrent major depressive disorder with atypical features	38694004
Recurrent major depressive episode NOS	191617002

Recurrent major depressive episodes, in full remission	191615005
Recurrent major depressive episodes, in partial or unspecified remission	191614009
Recurrent major depressive episodes, unspecified	191609005
Acute depression	712823008
Adjustment disorder with depressed mood	57194009
Adjustment disorder with mixed anxiety and depressed mood	782501005

Additional file 2 Variables used in the construction of the clinical pathways

One pathway indicates 45 mothers who later developed PPD had mental health disorders during pregnancy. Another pathway indicates mothers in both the PPD and non-PPD groups were prescribed [Opioids, Other Analgesics and Antipyretics, and Anti-inflammatory and Antirheumatic Products, non-steroids] prior to the refill of [Anti-inflammatory and Antirheumatic Products].



Additional file 3 Example of patterns identified from clinical sequences in the EHR

Variables
Anxiety history
Other disorder history
Antidepressants
Mood disorder history
Depression in pregnancy
Anxiety in pregnancy
Mental disorder in pregnancy
Palpitations

Diarrhea
Vomiting in pregnancy
Hypertensive disorder
Acute pharyngitis
Hemorrhage in early pregnancy antepartum
Threatened miscarriage
Abdominal pain
Migraine
Beta blocking agents
Antihistamines for systemic use
Hypothyroidism
Placental infarct
Single (vs. Married)
Deliveries by cesarean
Direct acting antivirals
Primigravida
Pre-eclampsia
Other antibacterials
ED visit
Abnormality of organs and/or soft tissues of pelvis affecting pregnancy
Diastolic blood pressure in the third trimester
False labor at or after 37 completed weeks of gestation
Race

Additional file 4 Characteristics of PPD cases across clusters

Variables (PPD=1)	**Cluster by PPD Risk**		
	High (N=1934, 6.72% PPD)	**Moderate (N=4129, 2.66% PPD)**	**Low (N=2886, 1.14% PPD)**
Demographics			
Age, mean (SD), year	34.68 (4.34)	33.85 (4.57)	34.82 (4.57)
Pre-pregnancy BMI, mean (SD), kg/m^2	24.11 (5.78)	22.71 (3.40)	24.68 (3.87)
Gestational Week, mean (SD), week	38.38 (2.74)	38.39 (2.42)	38.39 (3.61)
Race, No. (%)			
White	77 (59.23)	66 (60.0)	18 (54.55)
Asian	14 (10.77)	13 (11.82)	2 (6.06)
Black or African American	4 (3.08)	7 (6.36)	2 (6.06)
Other	16 (12.31)	15 (13.64)	4 (12.12)
Unknown	19 (14.62)	9 (8.18)	7 (21.21)
Marital Status, No. (%)			
Single (vs. Married)	27 (20.77)	19 (17.27)	7 (21.21)
ED Visits			
Pre-delivery, mean (SD)	1.17 (1.49)	0.95 (1.27)	0.76 (0.75)
Post-delivery, mean (SD)	0.15 (0.40)	0.08 (0.28)	0.12 (0.42)
Cesarean Section			
Yes, No. (%)	39 (30.0)	27 (24.55)	9 (27.27)

Additional file 5 Associations between PPD and the built environment variables in the study cohort

Variables	PPD	non-PPD	P-value
Number of bus stops within 500 m radius, mean (SD)	26.51 (10.12)	25.22 (10.0)	.04
Number of subway stations within 500 m radius, mean (SD)	1.88 (1.79)	1.81 (1.83)	.51
Parks Area within 500 m radius, mean (SD), m^2	433147.66 (660147.75)	464055.30 (660533.55)	.45
Bike Path Length within 500 m radius, mean (SD), m	30037.96 (14528.24)	29040.51 (15192.53)	.29
VKT of light vehicles within 500 m radius, mean (SD), 100,000 units	3179.51 (2335.33)	3287.15 (2240.08)	.44
VKT of heavy vehicles within 500 m radius, mean (SD), 10,000 units	3799.90 (2606.12)	3602.41 (2513.05)	.20
LUM index within 500 m radius, mean (SD)	0.64 (0.18)	0.64 (0.17)	.78
RetFar within 500 m radius, mean (SD)	0.25 (0.21)	0.24 (0.23)	.38
Number of Intersections within 500 m radius, mean (SD)	13.20 (8.50)	12.03 (7.74)	.03
Sidewalk Area within 500 m radius, mean (SD), 1000 m^2	915.62 (224.85)	907.52 (208.01)	.53
Ozone Concentration, mean (SD), $\mu g/m^3$	46.53 (0.44)	46.56 (0.50)	.28
$PM_{2.5}$ Concentration, mean (SD), $\mu g/m^3$	9.25 (0.46)	9.28 (0.47)	.25

Aim 2: Clinician Interviews

We conducted 30-minute interviews with 5 clinicians at Weill Cornell Medicine and Emory University, including one obstetric-gynecologist (Dr. Rochelle Joly), one reproductive psychiatrist (Dr. Alison Hermann), and three pediatricians specializing in allergy (Drs. Tricia Lee and Elizabeth Feuille) and pediatric obesity (Dr. Marianne Sharko). Interviews were conducted via teleconference in 2020 and 2021. Topics discussed include neighborhood safety, transportation options to parks and clinics, resources such as healthy food availability across neighborhoods and school districts, disparity due to payor status and inclination to self-disclosure of mental health conditions.

A central theme that arose during the interviews was the inability to fully leverage social determinants of health and environmental information, whether or not it is embedded in the electronic health records, during routine care. This is both due to the lack of established clinical evidence on the causal relationships between outcomes and the patients' residing environment, as well as clinicians' inability to take actions to intervene on social determinants of health. For example, clinicians can suggest healthy food options or provide education on the benefit of access to green space or cleaner air quality, but clinicians are often not able to directly provide such better environment for the patients. Instead, they may be able to adjust for therapeutic doses depending on patients home locations, although the effects may be limited. Clinicians commented that there are tools in the EHR and mobile applications to track environmental information such as air quality. However, frequently there is an information overflow and it is

unclear how to best pinpoint which information to use. Interviewed clinicians were cautious about the confounding factors in the relationship between the outcomes and neighborhood-level factors but agreed that the immediate environments, such as domestic partnership, housing quality, and food consumption, is more closely related to health outcomes than the broader built environment in the neighborhood. When time is available, they do ask about immediate environments as part of the screening to patients and care givers and consider them in the care.

When asked about using predictive models to help with clinical decision making, clinicians were hopeful that it may help guide patients to necessary care. They pointed out that it would be helpful if the risk prediction tool can be connected with suggested actions for the next steps. We think that this is an important takeaway from a methodological perspective of developing risk prediction tool for clinical applications. When asked about resources that may be helpful to have in clinics and embedded in EHR, clinicians suggested links to gym offering discounts, after school activities (and transportation offerings), farmers market locations, food pantry locations that clinicians can provide to patients.

Aim 3. Effect of the built environment on maternal mental health and children's allergic diseases

Abstract
Environmental factors have been associated with allergic diseases, and allergic diseases have also been associated with postpartum depression (PPD). A retrospective observational study was conducted using the EHR data regarding mothers and newborns in the New York City area to assess the effect of the built environment. Clinical data were extracted from the OMOP database for a five-year period between 2014 and 2019. Mothers and newborns were matched using the delivery date, date of birth as well as their demographic characteristics such as race and geography (i.e. FIPS code). Built environment data were extracted from multiple open data sources and matched with EHR data on geography. A univariate and multivariate analysis was conducted with mothers' PPD and children's allergy status as the outcomes and the environmental variables as predictors. This analysis included 152 mothers with PPD (cases) and 4,704 mothers without PPD (controls). Forty children were identified with allergies. The results indicate that mothers residing in neighborhoods with lower socioeconomic status are less likely to develop PPD. Environmental variables such as GINI inequality index and free blood pressure checks have a statistically significant and adverse effect on children with allergies.

Introduction
Postpartum or postnatal depression refers to major and minor episodes occurring within the first 12 months after delivery.[1] The prevalence of postpartum depression (PPD) is currently considered to be 10 to 15%. PPD is experienced in approximately 12% to 16% of women who give birth.[1] According to a study by Leung BM and Kaplan BJ, it is estimated that as many as 19% of new mothers may experience PPD within the first three months after their birth. Environmental factors have been associated with maternal mental health, and they can contribute to the increase of depression among women. These factors are low social capital, lack of social support and community "connectedness", ethnic segregation, and diversity and physical and social deprivation.[2]

PPD negatively affects the cognitive, social and developmental area in the lives of children. Maternal depression is strongly associated with poor infant growth, poor physical and emotional/behavior development, infant malnutrition and increased health problems.[1] PPD also contributes to congenital diseases in infants and failure to thrive medically. These illnesses of infants also increase the chance of PPD in new mothers.[1] To add, factors such as the mother's age, education and socioeconomic status can lead to less optimal childcare affecting the infant's cognitive status.[3]

Studies have shown that life events which cause stress on individuals can modify genes which later can influence the incidence of depression.[4] A study done by Mitchell et al. found that lower SES (education level), coupled with a reactive gene, had a significant effect on PPD, where mothers would have higher levels of PPD prevalent depending on the active allies and stressor types.[5] Other studies have found that demographics, such as age and education, have a negative impact on the prevalence of depression in new mothers.[6] Another study done by Shah et al. looked to understand if housing issues, rodent and bug infestation, had any relationship to depressive symptoms. They found that those who lived in homes with current roach infestation had almost three times the odds of experiencing high depressive symptoms, compared to groups who did not have an infestation in their homes.[7] These studies have

shown that there is a link between external environment factors and internal reactions related to the development of postpartum depression and depression symptoms severity.

There have been few studies that look to understand specific environmental factors such as, but not limited to, housing location, the availability of open space such as parks and heavy traffic causing noise pollution, having a negative effect on PPD in new mothers. We conducted a study to analyze if there are any sort of connections between the built environment and patients who have PPD, and ultimately to understand if there were any associations between PPD of mothers based on their environment, with the prevalence of infant allergy.

Methods

Clinical Data

A retrospective observational study was conducted using electronic health record (EHR) data regarding mothers and their newborns in New York City. The Observational Medical Outcomes Partnership (OMOP) Common Data Model database was used to extract mothers and infants from NewYork-Presbyterian Hospital (NYP). To identify and build cohorts of women with PPD and their infants who were inpatient and admitted from the emergency department between 2014 and 2019, Concept IDs including 9201 (Inpatient Visit) and 262 (Emergency Room + Inpatient Visit) were used. The ICD-10 code Z38 was used to identify infants born in NYP. The inclusion criteria for the study were mothers 16 years of age or older and infants younger than 1 year. After applying these criteria, 28,241 mothers and 27,395 infants were identified in the desired 5-year period. The year 2014 was excluded from the analysis as there were no mothers diagnosed with PPD in the data.

Because no direct link between delivering mothers and their newborns exists in the EHR data, multiple steps were followed to link them. The strategy applied used mothers' date of delivery to match with infants' date of birth. Because multiple mothers were still associated with multiple babies, a second layer of matching, based on geography (i.e. FIPS code), was conducted. Finally, the demographic characteristics such as race was used to further validate the match of each mother with each infant and filter the cohort. 8,700 mothers matched with infants, with the 152 mothers diagnosed with PPD designated as the cases for the study.

The study was further expanded to identify infants under five years of age who were diagnosed with common allergic diseases born to mothers with PPD. They were identified and extracted from the OMOP database using ICD-10 codes for anaphylaxis, food allergies, asthma, eczema/dermatitis and rhinitis. Forty infants were diagnosed with the allergic conditions. Univariate and multivariate analyses were conducted to identify the significance of environmental variables as predictors and children's allergic diseases as the outcome.

Environmental Data

To test the hypothesis, a database of factors representing the built environment and socioeconomic factors in New York City was created using open source data. Data were included if they had factors that could have a direct or indirect effect on maternal mental health and also had locational data available in the form of longitude and latitude for conversion to FIPS code. Exclusion criteria were unrelatedness to mental health and lack of locational data that could be converted to FIPS code (e.g. zip code or borough). These data were built on previous datasets, including FACETS and a dataset that expanded on it, VACCINe.[8] The sources of the built environment dataset included NYC Open Data, the American Community Survey, the USDA Food Access Research Atlas, the CDC's Agency for Toxic Substances and Disease Registry, and the EPA National Air Toxics Assessment.[9-13] The final list of variables included in the analysis can be grouped in the following categories: availability of health food, nature and transportation, availability of medical services, crime activity, income and unemployment, and additional socioeconomic variables. This built environment dataset was joined to the clinical dataset on the FIPS code.

To expand the possibilities for future research, a separate database with expanded data in the states of Connecticut, Pennsylvania, New Jersey, and New York was also curated. The variables in this expanded dataset include a subset of the NYC database variables as well as additional variables for Connecticut from Connecticut Open Data.[14] These additional variables for Connecticut include data related to food, SNAP benefits, parks and child care related services.

Statistical Analysis

A univariate analysis was conducted to test the relationship between environmental factors and PPD, with PPD status (PPD=152, no PPD=4,704) as outcome and environmental factors as predictors. A separate univariate analysis was conducted to test the relationship between environmental factors and children's allergies. A multivariate

analysis was also conducted using all thirty-two predictors to test the relationship with allergies. Further analysis including a Backward selection and ANOVA were conducted to determine the best model. Based on evidence of the relationship between the built environment and maternal health, the hypothesis was that a poor built environment is statistically correlated with the development of PPD in women and development of allergies in children.

Results

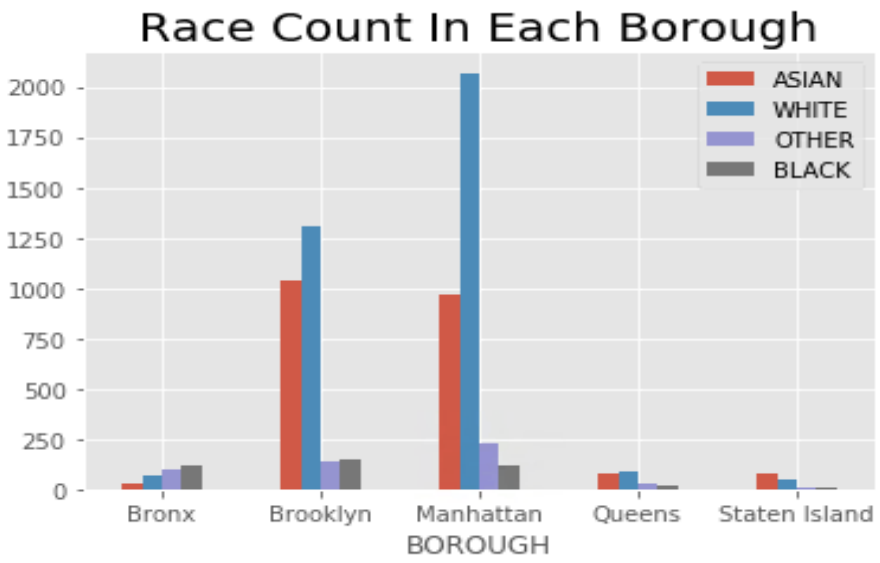
Effect of built environment on mother with PPD

The univariate analysis yielded seven variables with statistical significance. Table 1 displays the environmental variables that were significantly associated with mothers with PPD. They are socioeconomic status, household composition, minority status, social vulnerability, poverty rate, median household income, and unemployment rate. Median household income had an odds ratio of 1.05, which indicates that the odds of having PPD increases by 5% for each additional $10,000 increase in the median household income. The mothers who gave birth in NYP were mostly residents of Manhattan and Brooklyn and mostly White and Asian, as shown in Figure 1.

Table 1: Univariate analysis of mothers with PPD and environmental factors

Characteristic	**No PPD**, N = 4704	**PPD**, N = 152	**p-value**
Socioeconomic Status	0.36 (0.33)	0.23 (0.29)	<0.001
Household Composition	0.26 (0.27)	0.22 (0.25)	0.051
Minority Status	0.57 (0.23)	0.48 (0.20)	<0.001
Social Vulnerability	0.46 (0.30)	0.36 (0.26)	<0.001
Poverty Rate	0.15 (0.12)	0.11 (0.10)	<0.001
Median Household Income	88706 (47254)	99438 (38645)	<0.001
Unemployment Rate	0.06 (0.04)	0.05 (0.04)	0.004

Figure 1: Race distribution of mothers who delivered in NYP



Effect of built environment on children with allergies

Children born to mothers that were diagnosed with PPD were further studied to find the impact of the built environment on children's allergic diseases. A univariate analysis showed GINI index equality, free blood pressure checks, and number of flu vaccine locations are the significant predictors, with adverse effects on children with allergies. The GINI inequality index worked as a predictor with a significant adverse effect on children with allergies (p=0.02). The GINI inequality index had an odds ratio of 1.01, which indicates the odds of having allergic diseases increases by 1% for each point increase in the GINI. The multivariate analysis showed no significant impact from any of the thirty-two attributes containing the full model. A Backward selection was then performed using Akaike information criteria (AIC) step reduction which gave the predictors such as mother's race, misdemeanor count, violation rate, GINI, and borough. These predictors lead to meaningful reduction in AIC of the logistic model. An ANOVA with chi-square test showed that the full model is not significantly better than our trimmed down model obtained from backward selection. Thus, the predictors in the backward stepwise model are preferred.

Figure 2: The effect of environmental variables on mothers with PPD

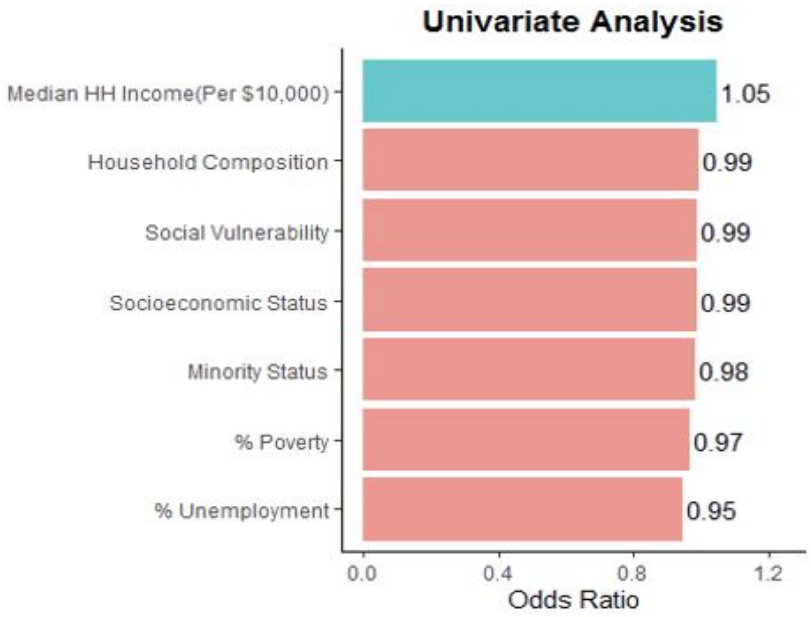
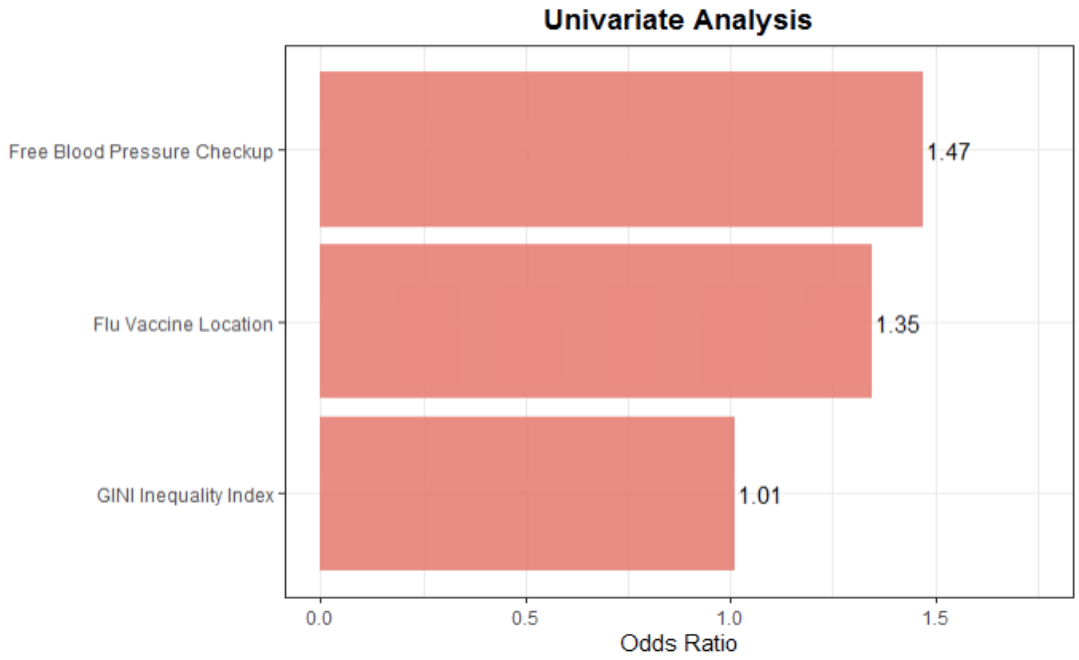


Table 2: Univariate analysis of children with allergies and environmental factors

Characteristic	Allergy, N = 40	No Allergy, N = 184	p-value
GINI Inequality Index	0.53 (0.05)	0.50 (0.06)	0.002
Free Blood Pressure Checkup			0.052
0	11 (28%)	78 (42%)	
1	11 (28%)	60 (33%)	
2	9 (22%)	30 (16%)	
3	9 (22%)	13 (7.1%)	
4	0 (0%)	2 (1.1%)	
5	0 (0%)	1 (0.5%)	
Flu Vaccine Location			0.2
0	14 (35%)	86 (47%)	
1	8 (20%)	46 (25%)	
2	10 (25%)	33 (18%)	
3	4 (10%)	13 (7.1%)	
4	4 (10%)	5 (2.7%)	
5	0 (0%)	1 (0.5%)	

Figure 3. The association between environmental variables as predictors and the allergic status of children as outcome



Discussion

Our study found that median household income is positively associated with PPD status, meaning more income in a house was related to an increased risk of PPD. Other variables were also found to be statistically significant in our analysis. Variables related to a lower socioeconomic status such as poverty levels and unemployment status were found to have a protective effect on PPD. Our hypothesis was rejected, as none of the built environmental factors collected had any effect on the prevalence of our PPD positive case group and our results indicated that higher rates of poverty are associated with a lower risk of PPD. Poverty rate was not a significant factor in the children with allergies either.

Dolbier et al. found that race did not have an effect on the prevalence of PPD, but rather subjective socioeconomic status (SES), a predictive factor of PPD.[15] A possible confounder in our study results could be that individuals from lower SES groups may not be reporting their PPD. This can occur due to several variables, such as cultural differences, education and resources available to mothers. A study conducted by Ahmed et al. also found that lower SES was associated with higher prevalence of PPD in mothers.[16] The protective effect of lower SES in our study may be an indication of a protective effect on the reporting itself. Further analysis and research must be conducted to evaluate any reporting bias to understand our patient population.

The majority of our patient population resides in the boroughs of Manhattan and Brooklyn and are White and Asian. According to the Census Bureau, Manhattan and Brooklyn are the third and first most populated boroughs within NYC, respectively.[17] The median income from 2018 for Manhattan was $82,459 with 15.6% of the population in poverty and median income for the same year in Brooklyn was $56,015 with 18.9% of the population in poverty, according to the Bureau.[17] It is not clear from the OMOP data if the mothers were screened for PPD or if the PPD was self-reported in our case group. It may be of benefit to learn whether patients from our sample who are in lower SES groups use resources such as WIC and SNAP benefits.

The adverse effect of certain variables such as the availability of free blood pressure checks and flu vaccines in children's allergies is harder to understand. One potential explanation is that these locations are placed specifically in areas with poor health, creating the illusion that these locations have an adverse effect on children's allergies.

Limitations

The environmental data collected were exclusively open source in nature and were somewhat limiting in completeness and in types of variables available. It is possible that not all FIPS codes of the patients are fully represented in all variables in the built environment dataset. The nature of PPD screening or self-reporting was not evident from the data. Data validation was required to make sure mothers' and infants' data were matched correctly to continue the study trajectory. The number of women diagnosed with PPD was small in our study. The population of children with allergies studied was also limited in our study. Broader collection of variable types may be needed to accurately understand the effects of the built environment on maternal mental health and children with allergic diseases.

Conclusion

This study analyzed the OMOP database to retrieve mothers who gave birth at NewYork-Presbyterian from 2014 to 2019. It looked at a small group of mothers with PPD to examine if there were any associations between environmental data and the incidence rate of PPD. The analysis was extended to PPD prevalence in mothers and the prevalence of allergies in children. Because median household income was positively associated with PPD status, our study rejected the hypothesis that a poor built environment is associated with PPD, though this may be limited by PPD reporting and open source data completeness. Environmental variables such as GINI inequality index and free blood pressure check locations were the only variables associated with children having allergies.

Future Direction

Additional research and data collection must be conducted to find possible associations between the built environment of mothers who develop postpartum depression and children who develop allergies, as our data were too limited to do so. Allergy data regarding current infants should be collected as mothers return for visits postpartum. Further collection of available data on both open source and clinical data related to the built environment is also necessary to fully understand what possible variables can have an effect on the incidence of postpartum depression within the patient population. The curated datasets for New York, New Jersey, Connecticut, and Pennsylvania can serve as the basis for future, more expansive research related to the built environment.

Reference

1. Leung BM, Kaplan BJ. Perinatal Depression: Prevalence, Risks, and the Nutrition Link—A Review of the Literature. J Am Diet Assoc.2009 Sep; 109(9) 1566-157.
2. Eastwood, J., Kemp, L. & Jalaludin, B. Explaining ecological clusters of maternal depression in South Western Sydney. BMC Pregnancy Childbirth 14, 47 (2014). https://doi.org/10.1186/1471-2393-14-47
3. Garcia Coll, C., Vohr, B. R., Hoffman, J., & Oh, W. (1986). Maternal and environmental factors affecting developmental outcome of infants of adolescent mothers. Journal of Developmental and Behavioral Pediatrics, 7(4), 230–236. https://doi.org/10.1097/00004703-198608000-00003
4. Caspi A, Sugden K, Moffitt TE, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science. 2003
5. Mitchell, Colter et al. “Role of mother's genes and environment in postpartum depression.” Proceedings of the National Academy of Sciences of the United States of America vol. 108,20 (2011)
6. Mayberry L, Horowitz J, Declercq E. Depression symptom prevalence and demographic risk factors among U.S. women during the first 2 years postpartum. Journal of Obstetric, Gynecologic, and Neonatal Nursing. 2007
7. Shah, Snehal N et al. “Housing Quality and Mental Health: the Association between Pest Infestation and Depressive Symptoms among Public Housing Residents.” Journal of urban health : bulletin of the New York Academy of Medicine vol. 95,5 (2018): 691-702. doi:10.1007/s11524-018-0298-7
8. Cantor MN, Chandras R, Pulgarin C. FACETS: using open data to measure community social determinants of health. J Am Med Inf Assoc. 2018;25(4):419–422.
9. City of New York. NYC Open Data. https://data.cityofnewyork.us. Accessed June 30, 2020.
10. United States Census Bureau. American Community Survey. https://www.census.gov/programs-surveys/acs/data.html. Accessed June 30, 2020.
11. US Department of Agriculture. Food Access Research Atlas. www.ers.usda.gov/data-products/food-access-research-atlas.aspx. Accessed June 30, 2020.
12. Agency for Toxic Substances and Disease Registry. The Social Vulnerability Index. http://svi.cdc.gov/. Accessed June 30, 2020.
13. US Environmental Protection Agency. National Air Toxics Assessment. https://www.epa.gov/national-air-toxics-assessment. Accessed June 30, 2020.
14. State of Connecticut. Connecticut Open Data. https://data.ct.gov/. Accessed August 4, 2020.
15. Dolbier, Christyn L et al. “Relationships of race and socioeconomic status to postpartum depressive symptoms in rural African American and non-Hispanic white women.” Maternal and child health journal vol. 17,7 (2013)
16. Ahmed HM, Alalaf SK, Al-Tawil NG. Screening for postpartum depression using

Kurdish version of Edinburgh postnatal depression scale. Arch Gynecol Obstet. 2012
17. U.S Census Bureau. Quick Facts NYC https://www.census.gov/quickfacts/fact/table/newyorkcountymanhattanboroughnewyork,bronxcountybronxboroughnewyork,queenscountyqueensboroughnewyork,kingscountybrooklynboroughnewyork,richmondcountystatenislandboroughnewyork,newyorkcitynewyork/HSG010219. Accessed August 30, 2020.

List of participating Authors by Aim

Aim 1.1

Yiye Zhang PhD MS[a]*†, Shuojia Wang MS[bc]†, Alison Hermann MD[a], Rochelle Joly, MD[a], Jyotishman Pathak PhD[a]

Weill Cornell Medicine, Cornell University, New York, NY, USA

b School of Public Health, Zhejiang University, HangZhou, Zhejiang, China

c Tencent Jarvis Lab, Shenzhen, China

† Both authors contributed equally to the manuscript

Aim 1.2

Yiye Zhang, PhD MS[1,2]; Mohammad Tayarani, PhD[3]; Shuojia Wang, PhD[4]; Yifan Liu, MS[1]; Mohit Sharma, MS[1]; Rochelle Joly, MD[5]; Arindam RoyChoudhury, PhD[1], Alison Hermann, MD[6]; Oliver H. Gao, PhD[7]; Jyotishman Pathak, PhD[1,6]

1. Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

2. Department of Emergency Medicine, Weill Cornell Medicine, New York, NY, USA

3. School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA

4. Tencent Jarvis Lab, Shenzhen Guangdong, China

5. Department of Obstetrics and Gynecology, Weill Cornell Medicine, New York, NY, USA

6. Department of Psychiatry, Weill Cornell Medicine, New York, NY, USA

Aim 2

Yiye Zhang, PhD MS
Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

Aim 3

Rimsha Azhar, MS, Yonaka Harris, MS, Afroza Sultana, MS, Mallika Viswanath, MS, Yiye Zhang, PhD MS
Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA