Womb for Concern: Air Quality Impact on Maternal Health Outcomes
Anja Vogt, Jade Glaister

Executive Summary

Project Description: We investigate the relationship between air quality index and maternal health outcome markers including infant birthweight and maternal eclampsia. The air quality data are from the EPA Outdoor Air Quality monitors, and maternal health data are Vital Statistics Natality Birth Data from the National Center for Health Statistics. We train classification models (Ridge and SVM) on natality data both with and without air quality index data to predict instances of maternal eclampsia and low infant birth weight. The models are evaluated using training and testing accuracies and rates of false negatives, which we aim to minimize. A highly accurate prediction model of this type can provide useful information to pregnant individuals about their risk level for eclampsia or infant low birth weight based on geographic, demographic, and health risk factors.

Data:
- Air Quality Data - Air Quality Data Collected at Outdoor Monitors Across the US | US EPA
- Maternal and Infant Health Data - Vital Statistics Natality Birth Data | NBER

Stakeholders

- Women and pregnant individuals
- Healthcare systems and reproductive health, child and infant health workers
- World aid organizations, international climate aid organizations
- Government, policymakers, and regulatory bodies interested in the social, political, and economic implications of air quality and climate
- Health technology, biotech, pharmaceutical companies

Key Performance Indicators (KPIs): Our main KPI is to minimize false negatives, such that the model minimizes the number of predicted normal birth weights and instances of no eclampsia that are in reality low birth weights or instances of eclampsia. We will also investigate training accuracies, testing accuracies, and recall for more information about the model's performance.

Model:

After preprocessing the data, we had two datasets: natality data not including air quality data, and natality data including air quality data. Each dataset was split into 80% training and 20% testing. Starting with the natality dataset without air quality data, we built a Ridge Classification model with balanced class weight and fit it to the training dataset with eclampsia as the outcome. Using this, we found training and testing scores, used the testing dataset to predict eclampsia outcomes, created confusion matrices, and evaluated the number of false positives. We repeated this for the natality dataset with air quality, then repeated this process with low birth weight as the outcome. For SVM Classification, we applied PCA and used 4 components (reaching a 95% approximation of the original data) for the model to reduce the intensity of the computations. We then repeated the same model building, fitting, and evaluation process as with Ridge Classification for eclampsia and low birth weight prediction, including and excluding air quality data.

Results

Solely using the number of false negatives to compare the models, the best model for predicting eclampsia would be SVM Classification without air quality data. However, the training and testing accuracies are significantly lower for this model, so our recommendation for predicting eclampsia is Ridge Classification with air quality Data, with 78.80% training and testing accuracy along with 471 false negatives. Using a similar evaluation, our recommendation for predicting Low Birth Weight is also Ridge Classification with air quality data, with 77.27% training accuracy and 77.29% testing accuracy along with 21274 false negatives. Adding air quality improved Ridge Classification models, but it worsened SVM classification models.

Future Goals:

It became challenging for our computers to run models working with such large datasets. Therefore, in the future, we will look into further methods of dimension reduction that are appropriate for the unique form of our training data.

In addition, we are interested in more thoroughly evaluating causal inference in the setting of maternal health outcomes, since there is a very large set of potential predictors we have access to, including air quality.

References

Cleveland Clinic. Eclampsia. Last modified March 28, 2023. Accessed June 27, 2025. https://my.clevelandclinic.org/health/diseases/24333-eclampsia.

Liu, Anqi, Hui Li, Qian Zhang, et al. "Machine Learning Models for Predicting Low Birth Weight: A Population-Based Retrospective Cohort Study." BMC Pregnancy and Childbirth 24 (2024): Article 20. https://pubmed.ncbi.nlm.nih.gov/40566361/.

National Center for Health Statistics (*2025*). Data File Documentations, *Natality*, *2004.* National Center for Health Statistics, Hyattsville, Maryland. https://www.nber.org/research/data/vital-statistics-natality-birth-data.

United States Environmental Protection Agency. *Outdoor Air Quality Data*. Accessed June 27, 2025. https://www.epa.gov/outdoor-air-quality-data

World Health Organization. Low Birth Weight. Nutrition Landscape Information System. Accessed June 27, 2025. https://www.who.int/data/nutrition/nlis/info/low-birth-weight.

Wu, Jun, et al. "Residential Exposure to Traffic and Spontaneous Preterm Birth." Environmental Health Perspectives 117, no. 9 (2009): 1409–1413. https://doi.org/10.1289/ehp.0800334.

Disclaimer: Any analyses, interpretations, or conclusions reached are credited to Jade Glaister and Anja Vogt and not to NCHS, which is responsible only for the initial data.