Associations between Air Pollution and Preterm Birth in the Bay Area

Deva Rani R. Dharmapalan

ABSTRACT

Air pollution has severe health implications, particularly for expectant mothers and their unborn babies. Using birth records from the California Department of Public Health and EPA data spanning 2018-2021, I investigated the association between air pollution levels (measured by AQI, $PM_{2.5}$, and O_3) and preterm birth rates across seven of the nine Bay Area counties. Employing linear regression analysis, I identified a significant positive correlation between air quality indicators such as AQI and O_3 and preterm birth rates. To delve deeper into the impact, I integrated racial demographic data from the 2020 US Census into the analysis. This addition revealed that regions with larger Black populations displayed the strongest correlation between air pollution and preterm births. This finding underscores the disproportionate burden of air pollution on marginalized communities, particularly Black populations. The results highlight the urgent need for enhanced air quality monitoring and targeted interventions in areas with high pollution levels, especially in communities with significant racial disparities. Implementing measures such as increased deployment of air quality monitors can aid in mitigating the adverse effects of air pollution on maternal and fetal health. By addressing these disparities, we can strive towards achieving healthier outcomes for all expectant mothers and their babies, regardless of race or socioeconomic status.

KEYWORDS

Air Quality, Environmental Health, Regression, Race, Gestation

INTRODUCTION

Pollution is becoming a growing health concern in the United States. Air pollution can lead to several diseases including asthma, lung disease, and cancer. (Terrell et al. 2022) Among those most vulnerable to the health repercussions of air pollution are expectant mothers and their developing fetuses. Extensive research underscores the detrimental effects of air pollution on birth outcomes. Exposure to fine particulate matter ($PM_{2.5}$) and ground-level ozone is associated with preterm birth (Bekkar et al. 2020), which can lead to further health complications in the child. Women exposed to elevated levels of NO_2 and benzene throughout their pregnancies have a higher risk for preterm birth (Llop et al. 2010). With disparities in air pollution exposure between neighborhoods, marginalized groups are potentially at higher risk of experiencing pollution-related preterm birth. (Liu et al. 2021)

Segregation and redlining have contributed heavily to disproportionate pollution exposure within between racial groups. Race is a better predictor than poverty in predicting exposure to pollution (Mikati et al. 2018). Black, Hispanic, Asian people, and people of color (as defined by the US Census) have higher exposure to $PM_{2.5}$ regardless of level of income, indicating systemic factors lead to disproportionate exposure within certain racial populations (Tessum et al. 2021). Black and Hispanic populations are more likely to live in areas burdened by polluting facilities (Mohai et al. 2009) and traffic related pollution sources such as highways and freeways. (Patel et al. 2021). Disparities in pollution exposure have contributed to the growing environmental injustice concern in the United States. With inadequate advocacy to protect communities of color from pollution exposure, these issues continue to pose threats to their health.

As of 2020, Black women experience 50% higher rates of preterm birth than white women. Overall, preterm birth rates among US-born non-white are higher than those of US-born whites. (Barreto et al. 2024). With lack of political representation and medical advocacy, non-white expectant mothers are at high risk of experiencing health risks due to air pollution. Black and Hispanic women have higher rates of preterm birth due to higher exposure to $PM_{2.5}$ and O_3. (Gray et al. 2014). The correlation between air pollution and preterm birth are higher within communities of color. Increasing black-carbon exposure from the 10th to the 90th percentile caused a 6.8 percent higher risk of preterm births among Black women. There is also a 2.1 percent higher risk among Hispanic women (Riddell et al. 2021). Coupled with racial discrimination, heavy air pollution in neighborhoods where non-white women reside may be contributing to higher rates of preterm births among these populations.

Air pollution has a notable impact on birth outcomes in the United States, and marginalized communities tend to suffer the most. The purpose of this study is to contribute to the ongoing research associating air pollution with preterm birth. The research focusing on race as a contributing variable in this association is limited. The Bay Area, a diverse area notorious for historical redlining, is an optimal study site for this research. (Nordone et al. 2020). While previous studies have been conducted in individual Bay Area cities, no studies have been conducted across the entire Bay Area.

To address this research gap, I conducted a study encompassing the nine Bay Area counties. I collected data on preterm birth rates and pollution levels across these counties. Employing a linear regression model, I calculated predicted preterm birth rates based on the pollution levels in each county. Additionally, by incorporating racial demographic information into our analysis, I examined the impact of race on the relationship between pollution and preterm birth rates.

METHODS

Study Site

I concentrated on the nine counties in the Bay Area (Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma). The Bay Area is a racially diverse area but has historically been subject to environmental justice issues. This is due to the historic placement of major freeways and highways in communities of color (Patel et al. 2021). Thus, communities of color in the Bay Area experience the burdens of pollution more so than white communities.

Data Collection

I began by collecting air pollution data from the EPA Pre-Generated Files, focusing specifically on AQI, $PM_{2.5}$ and O_3. The EPA data contains the median Air Quality Index (AQI) for each county for the years 2018-2021. AQI is typically measured in dimensionless units. It is used to quantify the quality of air in a particular location based on various pollutants such as particulate matter ($PM_{2.5}$ and PM_{10}), ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and carbon monoxide (CO). The pollutants $PM_{2.5}$ and O_3 are measured in micrograms per cubic meter of air. These measurements were separated by county and year. While AQI is measured in different units than $PM_{2.5}$ and O_3, and encompasses data from multiple pollutants, I will be referring to these three factors as "pollutants" henceforth.

Additionally, I needed data on preterm births within the Bay Area. I collected the preterm birth data from the California Department of Public Health, which is under the purview of the Center for Disease Control and Prevention (CDC). This dataset encompasses information on the rate of preterm birth from each county in the United States. The most recent data available is from 2021, so I used data spanning 2018-2021. Preterm birth is defined as births delivered after fewer than 37 weeks of gestation. Preterm birth rates shown in this data is the number of preterm births per 100 live births. Preterm birth rates are categorized by county and year. The dataset provides the upper and lower 95% confidence intervals for each preterm birth rate.

Lastly, I required racial demographic data. I collected US census data from 2020 which included racial breakdowns from each Bay Area County. The races are broken down into 6 categories (White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, and Two or more races). I excluded "Two or more races" and "Native Hawaiian and Other Pacific Islander" from my analysis. In the US Census, the classification of race is distinct from that of Hispanic origin. The data of each of these racial categories will be analyzed as percentages of the total population in each respective county.

Analysis

To determine whether the preterm birth data was viable, I visualized the rates of preterm birth within each county for each year along with its 95% confidence interval. It should be noted that Marin and Napa counties had wider confidence intervals than that of the other seven counties.

By running a regression analysis in Excel, I calculated the R-squared value of each pollution rate and the preterm birth rate of the respective year and county. Additionally, I ran an ANOVA test to assess the levels of variability within the regression model and p-values and t statistics were noted for each regression model. The R-squared values were then compared to find pollutants with particularly high correlations to preterm birth rate. Pollutants with a p-value less than 0.05 would be considered statistically significant. This approach allowed me to identify pollutants potentially linked to preterm birth within the Bay Area.

To determine whether race is a confounding factor in determining whether air pollution can increase the risk of preterm birth, I conducted a regression model of each race and pollutants that had shown a significant correlation to preterm birth rates. The races that increased the R-square value would be considered confounding variables. (Carlberg, 2016). This analysis would help elucidate the relationship between race, air pollution and preterm birth.

RESULTS

I found the highest rates of preterm birth rates in Contra Costa (9.2%), Solano (8.7%), and Santa Clara (8.65%) between 2018 and 2021. Figure 1 shows the preterm birth rate of each county along with the 95% confidence intervals given in the CDC data set. Marin and Napa have the widest confidence intervals and were thus excluded from the subsequent results to ensure accuracy. (Figure 1)

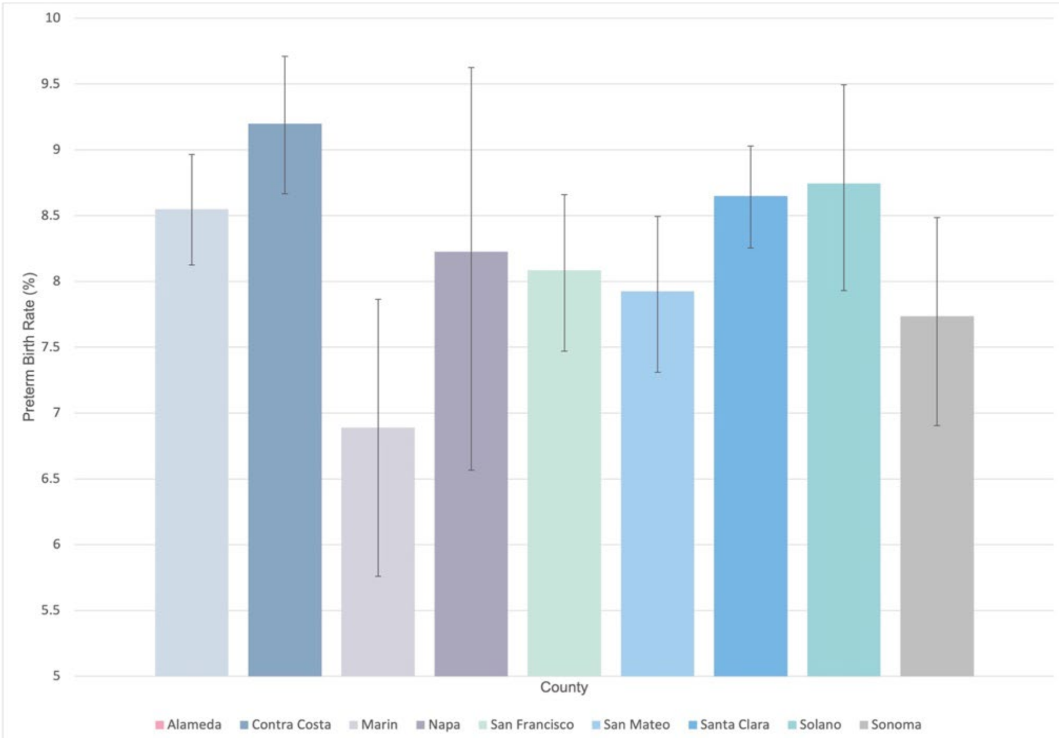

Figure 1. Preterm birth rates per county. Error bars display the 95% confidence interval. Data collected from the California Department of Public Health (2018-2021).

Pollution Rates

I calculated the median pollution rates across 2018-2021 for each pollutant. I observed the highest AQI rates in Santa Clara (AQI = 44) and Alameda (AQI = 44) counties. The highest rates of O_3 were observed in Santa Clara County (0.0365 ug/m^3). The highest rates of $PM_{2.5}$ were observed in Solano County (8.225 ug/m^3). The county with the second highest rates of $PM_{2.5}$ was Santa Clara County (7.7 ug/m^3). (Figure 2A-C)

Figure 2A. Median AQI rates per county. Data collected from EPA Pre-Generated Files (2018-2021).

Figure 2B. Median O_3 rates per county (ug/M^3). Data collected from EPA Pre-Generated Files (2018-2021).

Figure 2C. Median $PM_{2.5}$ rates per county (ug/M^3). Data collected from EPA Pre-Generated Files (2018-2021).

Air Pollution & Preterm Birth

The result of the regression models indicated that there was a positive correlation between air pollution and preterm birth rates. AQI was correlated with preterm birth and is statistically significant based on the p-value (R^2=0.304, p-value=0.002). O_3 was correlated with preterm birth (R^2=0.395) and is statistically significant (p-values = 3.42 E-4). There was a positive correlation between $PM_{2.5}$ and preterm birth rates (R^2=0.072, p-value = 0.167) but the p-value is too high to diagnose this relationship (Table 1).

Table 1: Regression Statistics (R^2, p-value, t stat) Correlating Air Pollutants and Preterm Birth

	Pollutant		
	AQI	$PM_{2.5}$	O_3
R^2	0.304	0.072	0.395
p-value	0.002	0.167	<0.001
T statistic	3.371	1.421	4.120

Race

The multiple regression model accounting for race revealed that adding Black demographics as another independent variable resulted in a greater correlation between AQI and preterm birth rates (R^2 = 0.373, p-value=0.04). Hispanic demographics also resulted in a greater correlation between AQI and preterm birth rates (R^2= 0.34, p-value=0.003). American Indian & Alaska Native, Asian, and White demographics also slightly increased R^2 values. All results are compared to AQI and preterm birth regression statistics. (Table 2)

Table 2: Regression Statistics (R^2, p-value, t stat) *correlating* Race, AQI and Preterm Birth

	Race					
	Black	AIAN	Asian	Hispanic	White	AQI Only
R^2	0.373	0.311	0.309	0.340	0.314	0.304
p-value	0.042	0.004	0.003	0.003	0.039	0.002
T statistic	2.144	3.214	3.267	3.322	2.180	3.371

The multiple regression model on O_3, preterm birth, and race displayed the largest increase in correlation when Black demographics were included (R^2= 0.512, p-value = 0.001). White demographics also increased the correlation (R^2= 0.483, p-value = 8.2E-4). American Indian & Alaska Native, Asian, and Hispanic demographics also slightly increased R^2 values. All results are compared to AQI and preterm birth regression statistics. (Table 3)

Table 3: Regression Statistics (R^2, p-value, t stat) Correlating Race, O_3 and Preterm Birth.

	Race					
	Black	AIAN	Asian	Hispanic	White	O_3 Only
R^2	0.512	0.421	0.416	0.406	0.483	0.395
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
T statistic	3.611	4.131	4.144	3.880	3.802	4.120

DISCUSSION

I investigated the association between air pollution and preterm birth within the Bay Area over the period spanning 2018 to 2021. Among the 7 counties studied, I found significant associations between median AQI and median O_3 and preterm birth rates. $PM_{2.5}$ data did not have significant associations with preterm birth rates. Correlations between AQI and preterm birth rates were higher within counties with a larger Black and Hispanic demographic. Correlations between O_3 and preterm birth rates were higher within counties with a larger Black and White population.

Air pollution and Preterm Birth

I found correlations between preterm birth and levels of AQI and O_3. R^2 values and p-values revealed a significant correlation between preterm birth and these two pollutants. This aligns with previous studies that had shown a correlation between O_3 and preterm birth (Ha et al. 2022). As AQI contains data from five major pollutants (ground-level ozone, particle pollution, SO_2, CO, NO_2), our results were consistent with our expectations. There was a positive correlation between $PM_{2.5}$ and preterm birth rates as well, however the p-value was too high to diagnose the relationship. This was an unexpected finding as many studies have shown a significant correlation between preterm birth and $PM_{2.5}$ exposure (Bekkar et al. 2020). As AQI data consists of all pollutant measurements, it can be concluded that air pollution overall is correlated with preterm birth, and a study focusing on separate pollutants may require more localized medical and pollution data. I found that O_3 was the pollutant with the highest correlation to preterm birth in the Bay Area. However, incorporating more pollutants into future studies will determine if other pollutants contribute more to preterm birth rates.

Race

I found that there is a higher rate of AQI-related preterm birth within counties with higher Black and Hispanic populations. Other racial categories also increased the R^2 value at lower rates. These two racial categories have been identified as being at a higher risk of pollution-related preterm birth in previous studies (Riddell et al. 2021). O_3-related preterm birth was higher within counties with higher Black populations. White populations were also found to increase this correlation slightly. This finding was inconsistent with previous studies that analyzed racial demographics. (Riddell et al. 2021). Generalization of racial demographics to 2020, rather than 2018-2021, may have contributed to the inaccuracy of our correlation statistics.

Limitations & Future Direction

I experienced some limitations due to lack of detailed data. While I would have ideally been working with medical data, which would include zip codes of pregnant mothers, I was unable to do so due to medical privacy protections. Preterm birth data from the years 2022 and 2023 were not available, so I was limited to data from previous years. Thus, results from this study may be outdated.

Future investigations should consider the impact of confounding variables such as diet, health care, and socioeconomics on the regression model. Expanding this study to a broader array of counties would provide a more comprehensive understanding of the relationship between air pollution, preterm birth, and racial demographics.

Implications

This study could inform future policy measures, preventing the placement of air pollution sources, such as major freeways, in communities that have been historically marginalized. This study underscores the necessity of low-cost air pollution monitors, which would be implemented throughout residential areas. These detectors can empower communities to assess their risk of developing health issues due to pollution and make informed decisions regarding their medical needs.

ACKNOWLEDGEMENTS

I extend my sincere gratitude to my instructors, Patina Mendez and Jessica Craigg, for their unwavering support and guidance throughout the writing journey. Special thanks to Ajay Pillarisetti for his invaluable contributions to the collection of air pollution data, showcasing his dedication to the field of Environmental Health. I am immensely grateful to my workgroup members for their insightful revisions and constructive feedback, which significantly enriched the quality of this thesis. My heartfelt appreciation goes to my family, whose encouragement has been a constant source of motivation. Specifically, I express my sincere appreciation for Jonathan, Jacqueline, Rebecca, Mooka, Smokey, and Bruno Dharmapalan. Their support has nourished my curiosity for interdisciplinary studies and a broad range of research topics. Thank you to all who have played a part in shaping this work and contributed to my academic growth.

REFERENCES

Barreto, A., B. Formanowski, M.-M. Peña, E. G. Salazar, S. C. Handley, H. H. Burris, R. Ortiz, S. A. Lorch, and D. Montoya-Williams. 2024. Preterm Birth Risk and Maternal Nativity, Ethnicity, and Race. JAMA Network Open 7:e243194.

Bekkar, B., S. Pacheco, R. Basu, and N. DeNicola. 2020. Association of Air Pollution and Heat Exposure With Preterm Birth, Low Birth Weight, and Stillbirth in the US: A Systematic Review. JAMA Network Open 3:e208243.

Carlberg, C. G. 2016. Regression analysis: Microsoft Excel. Que, Indianapolis, Indiana.

Gray, S. C., S. E. Edwards, B. D. Schultz, and M. L. Miranda. 2014. Assessing the impact of race, social factors and air pollution on birth outcomes: a population-based study. Environmental Health 13:4.

Ha, S., V. Martinez, and A. M. Chan-Golston. 2022. Air pollution and preterm birth: A time-stratified case-crossover study in the San Joaquin Valley of California. Paediatric and Perinatal Epidemiology 36:80–89.

Liu, J., L. P. Clark, M. J. Bechle, A. Hajat, S.-Y. Kim, A. L. Robinson, L. Sheppard, A. A. Szpiro, and J. D. Marshall. 2021. Disparities in Air Pollution Exposure in the United States by Race/Ethnicity and Income, 1990–2010. Environmental Health Perspectives 129:127005.

Llop, S., F. Ballester, M. Estarlich, A. Esplugues, M. Rebagliato, and C. Iñiguez. 2010. Preterm birth and exposure to air pollutants during pregnancy. Environmental Research 110:778–785.

Mikati, I., A. F. Benson, T. J. Luben, J. D. Sacks, and J. Richmond-Bryant. 2018. Disparities in Distribution of Particulate Matter Emission Sources by Race and Poverty Status. American Journal of Public Health 108:480–485.

Mohai, P., P. M. Lantz, J. Morenoff, J. S. House, and R. P. Mero. 2009. Racial and Socioeconomic Disparities in Residential Proximity to Polluting Industrial Facilities: Evidence From the Americans' Changing Lives Study. American Journal of Public Health 99:S649–S656.

Nardone, A., J. A. Casey, R. Morello-Frosch, M. Mujahid, J. R. Balmes, and N. Thakur. 2020. Associations between historical residential redlining and current age-adjusted rates of emergency department visits due to asthma across eight cities in California: an ecological study. The Lancet Planetary Health 4:e24–e31.

Patel, L., E. Friedman, S. A. Johannes, S. S. Lee, H. G. O'Brien, and S. E. Schear. 2021. Air pollution as a social and structural determinant of health. The Journal of Climate Change and Health 3:100035.

Riddell, C. A., D. E. Goin, R. Morello-Frosch, J. S. Apte, M. M. Glymour, J. M. Torres, and J. A. Casey. 2022. Hyper-localized measures of air pollution and risk of preterm birth in Oakland and San Jose, California. International Journal of Epidemiology 50:1875–1885.

Terrell, K. A., and G. St Julien. 2022. Air pollution is linked to higher cancer rates among black or impoverished communities in Louisiana. Environmental Research Letters 17:014033.

Tessum, C. W., D. A. Paolella, S. E. Chambliss, J. S. Apte, J. D. Hill, and J. D. Marshall. 2021. $PM_{2.5}$ polluters disproportionately and systemically affect people of color in the United States. Science Advances 7:eabf4491.