Effect of PM 2.5 on Asthma-related visits to the ER in Los Angeles

Henry Caldera, Eunice Ngai, Juliana Gómez Consuegra

Introduction and Data

Asthma affects citizens of all ages and genders. The financial cost of asthma exceeds $81 billion annually and is responsible for over 3,500 deaths annually in the United States[1]. Notably, in 2019-2020, Los Angeles County had a 15.1% prevalence of asthma, meaning 1.4 million residents are affected by asthma[2]. While there have been studies done to find associations of asthma prevalence and exacerbation with air pollution, analysing data from Los Angeles County could help identify the major cause of asthma in the county, thus providing direction for policymakers to improve the well-being of their constituents, particulary "low-income communities, communities of color, tribal nations and other disadvantaged groups who continue to be overburdened by unacceptable levels of pollution"[3] and children[4]. This study uses linear regression analysis to estimate the effect of air pollution on the rate of asthma in Los Angeles County.

The data comes from CalEnviroScreen 4.0, which was developed by the Environmental Health Hazard Assessment (OEHHA), an office of the California Environmental Protection Agency. CalEnviroScreen 4.0 is a model that incorporates 21 environmental, health, or socioeconomic variables for each census tract to "help identify Californian communities that are disproportionately burdened by multiple sources of pollution"[3]. The observational data used in the CalEnviroScreen 4.0 model is gathered from national and state sources, and includes variables such as: Ozone pollution, $PM_{2.5}$ pollution, toxic factory chemical release, vehicle emissions, level of education, level of poverty, level of employment, etc. The unit of observation is the 2343 census tracts in Los Angeles county. However, data is not available directly from each census tract, therefore pollution exposure variables, such as `PM2.5`, `Ozone`, `Traffic`, and `Diesel_PM`, were modeled by CalEnviroScreen based on data from sensors found throughout the county. Data for socioeconomic variables, such as `Housing_Burdened`, were collected from the American Community Survey (ACS). All exploration and model building were performed on a 30% sub-sample of the data. The remaining 70%, totaling 1582 rows, was used to generate the statistics in this report.

Operationalization of Key Concepts and Key Modeling Decisions

To operationalize air pollution, we made use of the modeled data by CalEnviroScreen. To estimate the concentration of `PM2.5`, the CalEnviroScreen used the weighted average, in $\mu g/m^3$, of measured monitor concentrations and satellite observations of particles that have a diameter of 2.5 μm or less, from 2015 to 2017. There exist a variety of ways to operationalize air pollution, but `PM2.5` was chosen because it encapsulates several air pollutants that are able to travel deeply into the lungs, even entering the bloodstream[5]. To operationalize the rate of asthma, we used the rate of Emergency Department (ED) visits for asthma-related

[1] Perez MF, Coutinho MT. An Overview of Health Disparities in Asthma. Yale J Biol Med. 2021 Sep 30;94(3):497-507. PMID: 34602887; PMCID: PMC8461584.

[2] California Asthma Dashboard. (2020). Environmental Health Investigations Branch. California Department of Public Health.

[3] Blumenfeld, J. Zeise, L. (2021) CalEnviroScreen 4.0 Report.

[4] McConnell R, et al. (2010). Childhood incident asthma and traffic-related air pollution at home and school. Environ Health Perspect 118(7):1021-6

[5] Chankaew, K., et al. (2022). Spatial Estimation of PM2.5 Exposure and its Association with Asthma Exacerbation: A Prospective Study in Thai Children. Annals of Global Health, 88(1), 15.

symptoms per 10,000 visits, `Asthma`. ED utilization does not capture the full burden of asthma in a community because not everyone with asthma requires emergency care or had access to healthcare; however, we were unable to find data on the prevalence of asthma per census tract in Los Angeles County. Using the rates of asthma-related ED visits is a reasonable proxy variable for the prevalence of asthma. Following a review of relevant literature, we selected additional input variables known to be associated with prevalence of asthma[6] [7] [8] [9]. These variables are: `Ozone` (Mean of the daily maximum between May and October of 8-hour ozone concentration in ppm, averaged from 2017 to 2019), `Traffic` (Sum of traffic volumes adjusted by road segment length, in vehicle- kilometers per hour, divided by total road length in kilometers, within 150 meters of the census tract, in 2017), `Diesel_PM` (Spatial distribution of gridded diesel PM emissions in tons/year from on-road and non-road sources in 2016), `Pesticides` (Total pounds of 132 selected active pesticide ingredients, filtered for hazard and volatility, used in production-agriculture per square mile, averaged from 2017 to 2019), `Poverty` (Percent of the population living below two times the federal poverty level, 5-year estimate from 2015 to 2019), `Low_Birth_Weight` (Percent of infants with low birth weight, averaged over 2009-2015), `Housing_Burdened` (Percent of households in a census tract that both make less than 80% of the HUD Area Median Family Income and pay more than 50% of their income in housing, averaged from 2013 to 2017).

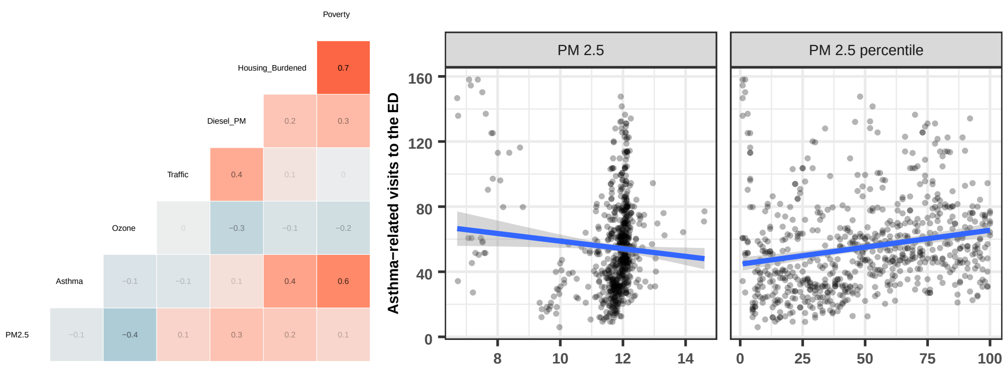

Figure 1: a. Correlations between all variables in the dataset b. Distribution of PM 2.5 data before (left panel) and after (right panel) transforming it into percentiles.

Exploratory data analysis revealed 81 census tracts with incomplete data, therefore we removed them (3.45% of the data). We found that the variable `Pesticides` was sparsely populated, with only 443 census tracts having information, so it was removed from the dataset. A correlation matrix (see Figure 1a) showed a high correlation of 0.7 between `Poverty` and `Housing-burdened`; `Poverty` was removed in favor of `Housing-burdened` because it encapsulates a broader scope of socioeconomic factors. The variable `Low_Birth_Weight` was excluded from our model because "previous studies investigating criteria pollutant data, ambient carbon monoxide, nitric oxide, nitrogen dioxide, nitrogen oxides, PM10, and PM2.5 exposures were linked with increased odds of term low birth weight for third-trimester and entire-pregnancy averages"[10], making it an output variable among the covariates. Our final dataset with complete information for all covariates included 2262 census tracts. We transformed our data using percentile ranks, for two reasons. Firstly, the researchers who built the CalEnviroScreen 4.0 model advise transforming the covariates to percentile rank before doing statistical analysis because the covariates in their model have "varying underlying distributions"[3]. Further, scatterplot visualization of each covariate with `Asthma` revealed high-density clustering of datapoints for `PM2.5` (Figure 1b) and medium-density clustering for `Traffic` and `Diesel_PM`, therefore the percentile rank transformation has the additional benefit of widening the covariate distributions (see Figure 1b).

[6] Last JA, et al. (2017). Ozone and Oxidant Toxicity. In: Respiratory Toxicology. Elsevier Inc., pp. 389-402.
[7] Patel MM, et. al (2013). Traffic-related air pollutants and exhaled markers of airway inflammation and oxidative stress in New York City adolescents. Environ Res 121:71-8.
[8] Hernández, A. F., et al. (2011). Pesticides and asthma. Current opinion in allergy and clinical immunology, 11(2), 90–96.
[9] Meltzer R, Schwartz A (2016). Housing affordability and health: evidence from New York City. Housing Policy Debate 26(1):80-104.
[10] Ghosh et al.,(2012). Assessing the influence of traffic-related air pollution on risk of term low birth weight on the basis of land-use-based regression models and measures of air toxics. American journal of epidemiology, 175(12), 1262–1274

Results and Limitations

In our first regression model, we included only the covariate `PM2.5` because it was the variable that initiated our research and we sought to investigate its effect on `Asthma`. Given the wide scope of factors that affect `Asthma`, ranging from environmental to socioeconomic, the first model had very poor goodness of fit and the coefficient of `PM2.5` was inflated to compensate for omitted confounding variables. Our second regression model attempted to correct for omitted confounding environmental factors with the inclusion of `Diesel_PM`, `Ozone`, and `Traffic`, in addition to `PM2.5`. Our third and final regression model attempted to correct for omitted confounding socioeconomic factors with the inclusion of `Housing_Burdened`, in addition to the 4 environmental variables:

$$\widehat{Asthma\_ED}\_visits = \beta_0 + \beta_1 \, \texttt{PM2.5} + \beta_2 \, \texttt{Diesel\_PM} + \beta_3 \, \texttt{Ozone} + \beta_4 \, \texttt{Traffic} + \beta_5 \, \texttt{Housing\_Burdened}$$

Table 1: Regression Results

	Dependent variable:		
	Number of visits to the ED due to Asthma		
	(1)	(2)	(3)
Constant	45.78***	55.55***	42.68***
	(1.60)	(2.88)	(2.63)
Ozone		−0.08***	−0.09***
		(0.03)	(0.02)
Traffic		−0.11***	−0.07***
		(0.03)	(0.02)
‘PM 2.5‘	0.17***	0.15***	0.08***
	(0.03)	(0.03)	(0.02)
‘Diesel PM‘		0.02	−0.07**
		(0.03)	(0.03)
‘Housing Burdened‘			0.38***
			(0.02)
Observations	1,582	1,582	1,582
R^2	0.03	0.05	0.19
Residual Std. Error	27.22 (df = 1580)	26.97 (df = 1577)	24.87 (df = 1576)
F Statistic	49.99*** (df = 1; 1580)	20.87*** (df = 4; 1577)	75.25*** (df = 5; 1576)
Note:			*p<0.1; **p<0.05; ***p<0.01
			Robust standard errors in parentheses

Across all three models, the key coefficients of `PM2.5` were highly statistically significant, point estimates ranging from 0.08 to 0.17. Considering a hypothetical census tract in Los Angeles, by applying model (3), the rate of ED visits for asthma-related symptoms per 1,000,000 visits is predicted to increase by 8 if the level of `PM2.5` is increased by one percentile rank. Similarly, the `Housing_Burdened` coefficient in model (3) was 0.38. If the housing burden of a census track is increased by one percentile rank, the rate of ED visits for asthma-related symptoms per 1,000,000 visits is predicted to increase by 38. The goodness of fit for the 3 models increased from model (1) with $R^2 = 0.03$ to model (3) with $R^2 = 0.19$, which means our best-performing model was able to predict roughly 20% of outcomes. Contrary to the existing research findings, the coefficients for the air pollutant indicators `Diesel_PM`, `Ozone`, and `Traffic` were all negative in our models, indicating the number of asthma-related visits to the ED would decrease should the percentiles of each of these variables increase. Given the negative coefficients, we considered our models are not robust enough to provide actionable policy recommendations to reduce asthma-related visits to the ED from air pollution.

OLS regression requires two assumptions to be met. First, the data should be independent and identically distributed. Our data may not be independent, since census tracts are found close next to one another, so that pollution from one census tract can spillover to the next one. However, since it's a census of all locations, there should be no clustering issues [11]. The data is also not identically distributed, because the data for each variable was collected from different time periods, as detailed in the "Operationalization of Key Concepts" section. To overcome this issue, output variables were transformed into percentiles so that no assumptions about the underlying distributions would have to be made.

The second assumption is that a unique best linear predictor should exist. This occurs when none of the variables have finite variance, and when there is no multicolinearity. The distributions of the variables did not reveal heavy tails, except for `Traffic`. However, because there is a limit to the density of vehicles that can occupy the transportation infrastructure, the variance must be finite. Since `Poverty` and `Housing_Burdened` were highly-correlated (0.7), `Poverty` was dropped from our model to resolve issues of multicolinearity.

The low R^2 for our model suggests we may have omitted variable bias. We were not able to include data for several known causes of asthma in our model, such as genetics [12], diet [13], atopic sensitization [14], obesity [15], exposure to microbial agents [16], among others, as data is not available at a census tract level for Los Angeles County. The omission of data for dust mites from the model may be biasing the coefficients for traffic and ozone away from zero, as these should be positive, according to the literature. We do not suspect having issues of reverse causality or outcome variables in the right-hand side, since the only variable that could have a reverse causality relationship with asthma, `Low_Birth_Weight`, was removed from our model. We recommend that future studies should include collections of data at a census tract level, in order to better understand the effect of these variables on asthma.

Conclusion

This study sought to estimate the effect of particulate matter on asthma-related visits to Emergency Departments in Los Angeles County. The models suggest that every percentile rank increase in PM_2.5 leads to an additional 8-17 asthma-related visits to the ED per 1,000,000 visits to the ED. Remarkably, the largest impact on the model was seen when the socioeconomic variable `Housing_Burdened` was added. This suggests that residents of Los Angeles who experience higher poverty and higher cost-of-living expenses are more likely to have asthma-related visits to the ED. Future research should add more socioeconomic variables to the CalEnviroScreen model, since the goodness-of-fit of the models we tested was highest when a single socioeconomic variable was introduced. Given that CalEnviroScreen is "used by CalEPA and its boards and departments to aid in administering environmental justice grants, promote compliance with environmental laws, prioritize site-cleanup activities and identify opportunities for sustainable economic development and [...] identify disadvantaged communities in California"[3], we suggest that increasing the number of sensors around the county and including more socioeconomic variables in their model will add more granularity to CalEnviroScreen, allowing us in turn to build better regression models and to generate more precise and generalizable conclusions, which will most-benefit vulnerable populations.

---

[11] Kleemann (2023). pers. comm.

[12] Gore, C., and A. Custovic. "Can we prevent allergy?." Allergy 59.2 (2004): 151-161.

[13] Mihrshahi, Seema, et al. "The childhood asthma prevention study (CAPS): design and research protocol of a randomized trial for the primary prevention of asthma." Controlled clinical trials 22.3 (2001): 333-354

[14] Leung, Roland, et al. "Sensitization to inhaled allergens as a risk factor for asthma and allergic diseases in Chinese population." Journal of allergy and clinical immunology 99.5 (1997): 594-599.

[15] Redd, Stephen C. "Asthma in the United States: burden and current theories." Environmental health perspectives 110.suppl 4 (2002): 557-560.

[16] Des Roches, Anne, et al. "Immunotherapy with a standardized Dermatophagoides pteronyssinus extract. VI. Specific immunotherapy prevents the onset of new sensitizations in children." Journal of allergy and clinical immunology 99.4 (1997): 450-453.