# Lead Exposure: Exploring the Relationship between Blood Lead Levels and Incarceration Rates

A Socio-Spatial Analysis

2025-06-18

**Abstract**

Exposure to lead can have a wide variety of negative effects on a child's brain development and behavior, which may permanently affect learning and achievement outcomes. Previous research has demonstrated a relationship between lead exposure and damaging long-term behavior that leads to negative outcomes. Using data from 9,114 California census tract, this study examines the relationship between lead exposure and incarceration rates utilizing the blood lead levels (BLL) of children reported through state Medicaid enrollment. Using 2020 data from the Census Bureau, the Prison Policy Initiative, and the California Department of Public Health, we employed spatial cross-validation and multiple linear regression models to predict incarceration rates, considering demographic factors like income, race, and age. The analysis found weak correlations between BLL and incarceration rates census tract, with median income, median age, and racial composition of the census tract emerging as more significant predictors. The final model demonstrated moderate explanatory power (R-squared = 0.497). Our findings suggest that BLL is not a significant predictor of incarceration rates, emphasizing the stronger influence of socioeconomic and racial factors.

## 1 Introduction

In the mid-1990s, the US experienced a sharp downward trend in crime. Although determinants of crime are intertwined with many complex social and environmental factors, some researchers cite decreasing childhood lead exposure, accomplished through policies banning lead in paint and gasoline, as a plausible explanation (Wallace, 2017). A substantial body of research documents the adverse effects of lead exposure on brain development–children exposed to lead experience speech delays, learning disabilities, struggle with reading comprehension and math, and may experience behavioral changes (Schneider, 2023). Further, research by Talayero et al. (2023)[1] has highlighted a strong association between lead exposure during childhood and criminal tendencies during adulthood. Understanding and documenting evidence for adverse outcomes is particularly important given children from low-income households are more likely to be at risk for lead exposure (CDC, 2024). This research topic stands to contribute to important discussions at the intersection of crime, environmental racism, and social determinants of health.

Within this context, the authors constructed a study to determine whether the percentage of children tested above a safe threshold for blood lead level in a given census tract was a significant predictor of incarceration rate in that census tract. Using other demographic variables, like median age, median income, racial composition, and gender composition, our analysis sought to determine: **(1) Is there an association between percentage of children with a high BLL and incarceration rate** and **(2) If there is an association, is it significant compared to well-established predictors of likelihood to be incarcerated, like SES, race, gender, and age?** Using multiple linear regression models and spatial cross validation, we found that though a higher BLL was correlated with a higher incarceration rate in a given census tract, it was not a significant predictor of incarceration rate. Income in particular was a far stronger predictor of incarceration rate and was associated with the biggest decrease in odds of incarceration. These results highlight known disparities in the criminal justice system, but are limited by the intricate combination of social and institutional factors that increases one's likelihood of incarceration and the inability of this study to establish causal relationships between lead exposure and incarceration rate. The authors also note that blood lead levels as measured in children likely underestimates true lead exposure in a given census tract.

## 2 Materials & Methods

### 2.1 Data sets

Data for this study were obtained from three sources. California incarceration rates comes from a collaborative project by the Essie Justice Group and the Prison Policy Initiative (https://www.prisonpolicy.org/origin/ca/2020/tract.html). PPI reports incarceration rate by census tract via data collected by the US Census Bureau in 2020. This data is publicly available because of a 2020 law in California reforming prison gerrymandering. Each census tract in the incarceration rate data was matched with median age, percent male, median income, and racial composition data from the 2020 US Census. Finally, BLL data was obtained from the California Department of Public Health (CDPH). Notably, this data was aggregated from 2018-2022 to protect individual privacy. The final merged data set contained 9,114 California census tracts and their respective statistics relating to blood lead levels, income, incarceration rates, and racial demographics. The 94 census tracts that screened zero children for blood lead level were excluded from this analysis. Observations deemed false positives or false negatives by the CDPH were preemptively omitted by CDPH reporting procedures.

**Note:** Figures referenced throughout the report can be found in the Appendix section.

### 2.2 Data processing and analysis

This study investigated the relationship between incarceration rate and blood lead level while controlling for additional variables, including age, gender, race, and income. All variables were continuous. A new variable, *POC_other*, was created to account for the racial composition of the census tract by the percentage of the population that is Black, Native, or Hispanic non-white (Table A1).

Spatial correlation among nearby census tracts was likely. Tracts next to each other may share similar characteristics such as policing practices, built environments, and school systems that could potentially impact rates of incarceration, particularly near Los Angeles county (Figure B4). To address this independence problem, the authors employed spatial cross validation to attempt to mitigate the effects of spacial correlation. Figure C1 illustrate the division of census tracts into five spatial blocks, each representing a fold used in spatial cross-validation. These blocks were generated with a block range of 100 km, ensuring that observations within each fold are spatially clustered and that training and testing sets are geographically separated. Geographical distance is the only factor involved in generating the folds–other variables and their coefficients are not involved (Valavi et al., 2018).

## 3 Results

### 3.1 Descriptive Analysis and Initial Modelling

Descriptive analysis indicated that incarceration rates and income distributions within California census tracts were right skewed (Figures B1 & B2). The authors elected to log transform these variables, which produced a residual plot that appeared to satisfy conditions of constant variance and linearity. Normality was also satisfied (n > 30). No data points displayed a cook's distance greater than 0.5, and as a result, no points were deemed influential.

The highly correlated nature of race and income lead the authors to examine a potential interaction effect. In Figure B3, a categorical variable flagged *POC_other* values that were above and below the median *POC_other* value to create a categorical split. The relationship between imprisonment rate and income appeared more negatively correlated when the percentage of *POC_other* in the census tract was above the median, leading to the inclusion of an interaction term between race and income in the initial model.

### 3.3 Multiple Linear Regression Models

The multiple linear regression models revealed the significance of BLL in predicting incarceration rate, the effects of other examined variables, and the significance of the interaction term between *median_income* and *POC_other* (Table C2). The model was trained on bins 1-4 of the spatially cross validated data, which represented approximately 80% of the census tracts. *log_imprisonment_rt, log_med_income, perc_bll_indicator, POC_other, median_age, perc_male*, and the interaction between *POC_other* and *log_med_income* were included in the full model. Stepwise selection using AIC resulted in the exclusion of perc_bll_indicator, which the authors forced back into the model because it was the variable of interest. log_median_income (centered) demonstrated the strongest negative relationship with predicted incarceration rate ($\beta = 0.515$, $p \approx 0$). Increasing POC_other also increased the predicted rate of incarceration ($\beta = 1.016$, $p \approx 0$). The interaction term between income and race was significant ($\beta = 1.004$, $p \approx 0$). More men in a census tract, Perc_male, increased the predicted rate of incarceration ($\beta = 1.008$, $p \approx 0$). Blood lead level was not a significant predictor ($p > 0.005$). Finally, an increasing median age was associated with an increase in the predicted incarceration rate ($\beta = 1.005$, $p \approx 0$).

Testing the model on the 5th bin of the spatially cross validated data, which represented approximately 20% of the census tracts, yielded an R squared value of 0.497 and an RMSE of 190.679 (Table C4). VIF values for all variables were below 3 when interaction terms were removed.

## 4 Discussion

This analysis aligns with prior research on social determinants of crime, as race, income, and gender well-established factors influencing likelihood of incarceration (Ghandnoosh, 2023). One suprising exception was age. Liklihood of committing a crime peaks just after adolescence and rapidly diminishes thereafter (Shulman, Steinberg, and Piquero, 2013). This analysis found that predicted incarceration rate *increases* with an increasing median_age. This is likely a limitation of using *median_age,* as the age–related relationship with incarceration is specific to the percentage of the population that is around 18-25 years old.

Notably, BLL is not statistically significant in predicting incarceration rate. The outstanding significance of other established factors, such as income, reveals that differing results may be due to the misunderstood or overstated effect of lead exposure on incarceration rate. At the very least, this analysis indicates that BLL is not a useful metric on its own. However, the R squared value of 0.497 indicates that the model is fairly successful at capturing complex real-world drivers of incarceration rates.

An important limitation of our current analysis is the method of measuring BLL in a given census tract. BLL levels in children were used to approximate to approximate risk of lead exposure for all people in the census tract, which may not be an accurate representation of the reality. A better method would be to combine BLL testing in children with other forms of measuring potential for lead exposure, like water testing or neighborhood environmental assessments. Future research should strive to confirm the lack of influence of BLL on incarceration rates through examining other geographical regions or measurements of lead exposure.

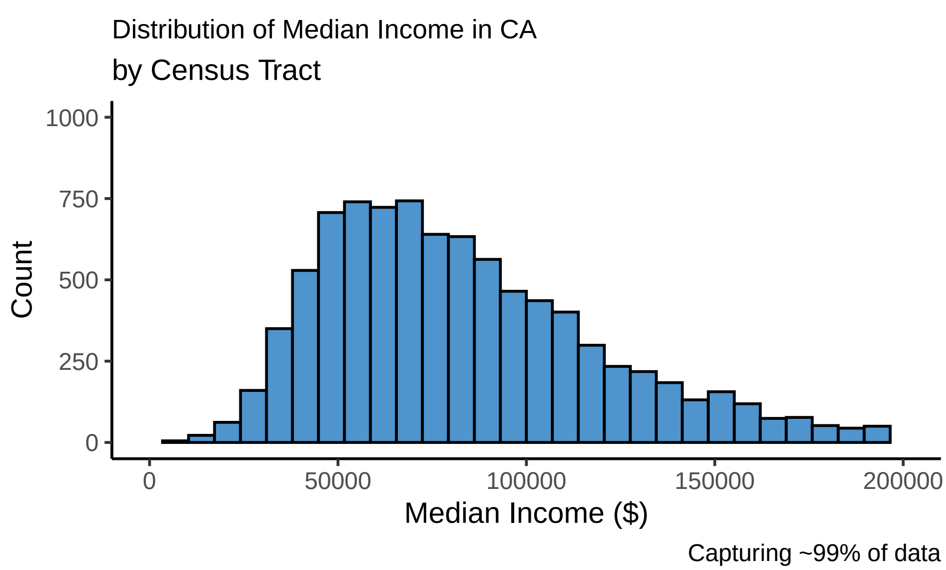
# 5 Appendix

## 5.1 Appendix A - bll_data Dictionary (variables in processed dataset)

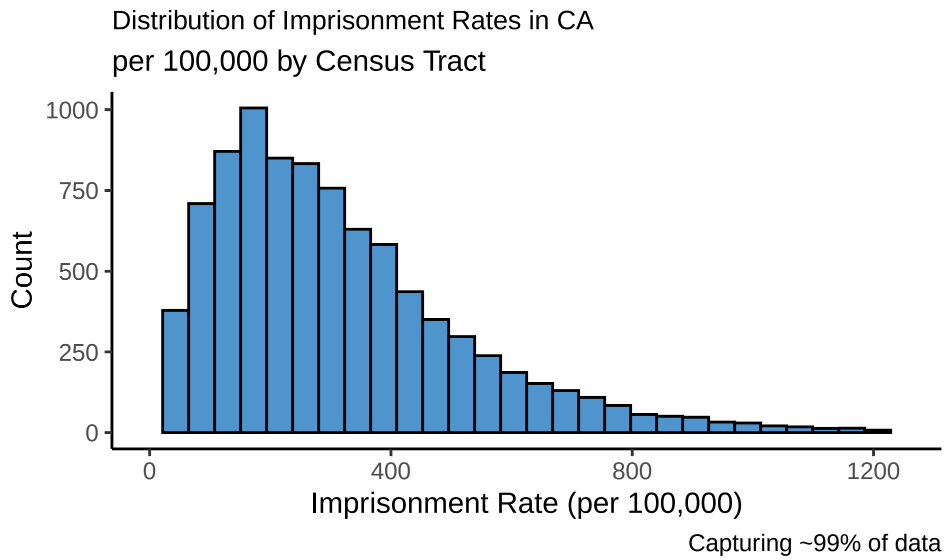
| Variable | Description |
|---|---|
| census_tract | California census tract (categorical) |
| city | California city (categorical) |
| num_bll | Total number of children in a census tract whose blood was tested for lead (numerical) |
| num_bll_indicator | The number of tested children in a census tract under 6 that have a blood lead level of 3.5mg or greater (numerical) |
| perc_bll_indicator | The percentage of tested children in a census tract under 6 that have a blood lead level of 3.5mg or greater (numerical) |
| POC_other | The percentage of a census tract population that is Black, Hispanic, or Native American |
| num_prison | The number of imprisoned people in a census tract (numerical) |
| total_pop_2020 | Population of census tract in 2020 (numerical) |
| imprisonment_rt | Imprisonment rate per 100,000 people in a census tract (numerical) |
| median_age | Median age in census tract (numerical) |
| perc_male | Proportion of males in census tract (%, numerical) |
| med_income | Median income of census tract ($, numerical) |

**(Table A1):** Variable Definitions

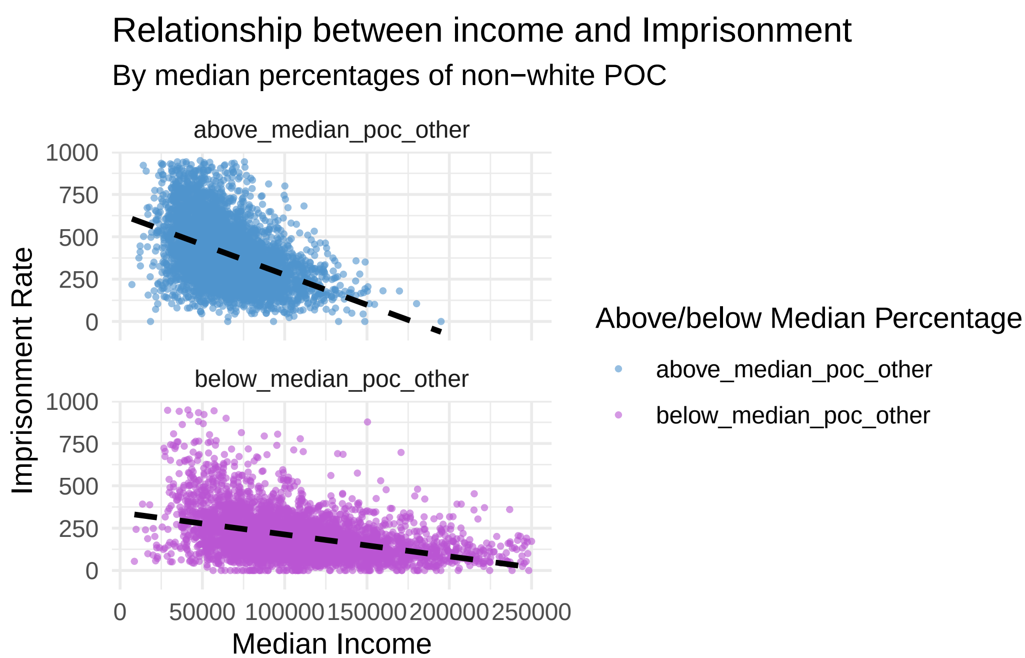
## 5.2 Appendix B - Descriptive Analysis



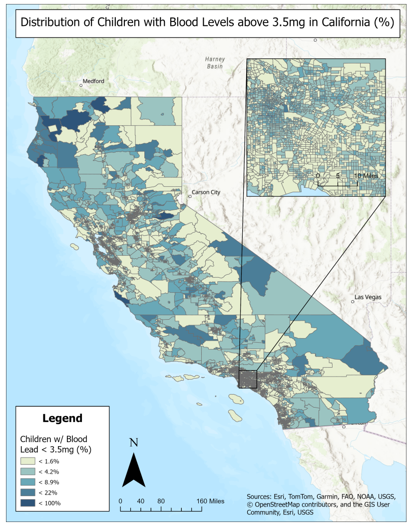

**(Figure B1):** Income distribution in California census tracts




**(Figure B2):** Incarceration rate distribution in California census tracts

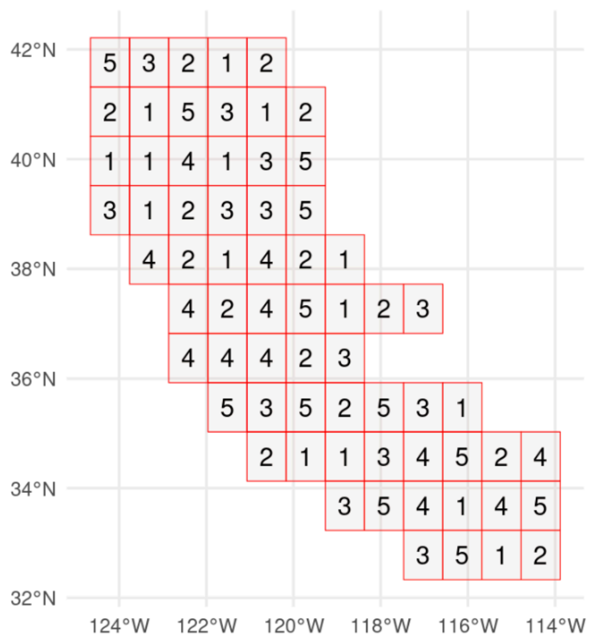



**(Figure B3):** Interaction effects between income and race

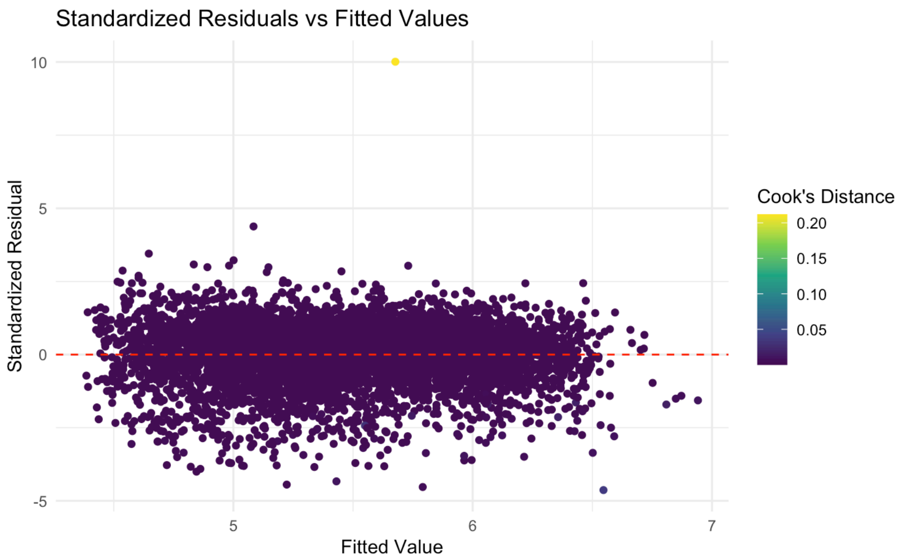



**(Figure B4):** Spatial distribution of census tracts in California with BLL

### 5.3 Appendix C- Spatial CV and Multiple Linear Regression Results




**(Figure C1):** Data divided into five spatial blocks, each representing a fold used in spatial cross-validation.




**(Figure C2):** Condition Checking for MLR

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 4.496 | 0.148 | 30.358 | 0.000 |
| log_med_income_c | -0.663 | 0.026 | -25.624 | 0.000 |
| perc_bll_indicator | -0.003 | 0.003 | -1.045 | 0.296 |
| POC_other | 0.016 | 0.001 | 25.557 | 0.000 |
| median_age | 0.005 | 0.001 | 4.476 | 0.000 |
| perc_male | 0.008 | 0.003 | 2.915 | 0.004 |
| log_med_income_c:POC_other | 0.004 | 0.001 | 4.522 | 0.000 |

**(Table C3):** Final Model Output

| .metric | .estimator | .estimate |
|---|---|---|
| rmse | standard | 190.679 |
| rsq | standard | 0.497 |

**(Table C4):** Evaluating Model Performance

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 4.889 | 0.013 | 380.335 | 0 |
| POC_other | 0.023 | 0.000 | 55.156 | 0 |

## 6 References


Ghandnoosh, N. (2023, October). One in five: Ending racial inequity in incarceration. The Sentencing Project. https://www.sentencingproject.org/reports/one-in-five-ending-racial-inequity-in-incarceration/

Princeton School of Public and International Affairs. (n.d.). Decrease in lead exposure in early childhood may be responsible for drop in crime rate. https://spia.princeton.edu/news/decrease-lead-exposure-early-childhood-may-be-responsible-drop-crime-rate

Schneider, J. S. (2023). Neurotoxicity and outcomes from developmental lead exposure: Persistent or permanent? Environmental Health Perspectives, 131(8), 085002. https://doi.org/10.1289/EHP12371

Shulman, E. P., Steinberg, L. D., & Piquero, A. R. (2013). The age-crime curve in adolescence and early adulthood is not due to age differences in economic status. Journal of Youth and Adolescence, 42(6), 848–860. https://doi.org/10.1007/s10964-013-9950-4

Talayero, M. J., Robbins, C. R., Smith, E. R., & Santos-Burgoa, C. (2023). The association between lead exposure and crime: A systematic review. PLOS Global Public Health, 3(8), e0002177. https://doi.org/10.1371/journal.pgph.0002177

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution, 10(2), 225–232. https://doi.org/10.1111/2041-210X.13107