Published in final edited form as:
Atmos Environ (1994). 2024 April 01; 322: . doi:10.1016/j.atmosenv.2024.120374.

Exposure to particulate matter and ozone, locations of regulatory monitors, and sociodemographic disparities in the city of Rio de Janeiro: Based on local air pollution estimates generated from machine learning models☆

Honghyok Kim[a,*], **Ji-Young Son**[b], **Washington Junger**[c], **Michelle L. Bell**[b]

[a]Division of Environmental and Occupational Health Sciences, School of Public Health, University of Illinois Chicago, Chicago, IL, United States

[b]School of the Environment, Yale University, New Haven, CT, United States

[c]Institute of Social Medicine, State University of Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

South America is underrepresented in research on air pollution exposure disparities by sociodemographic factors, although such disparities have been observed in other parts of the world. We investigated whether exposure to and information about air pollution differs by sociodemographic factors in the city of Rio de Janeiro, the second most populous city in Brazil with dense urban areas, for 2012–2017. We developed machine learning-based models to estimate daily levels of O_3, PM_{10}, and $PM_{2.5}$ using high-dimensional datasets from satellite remote sensing, atmospheric and land variables, and land use information. Cross-validations demonstrated good agreement between the estimated levels and measurements from ground-based monitoring stations: overall R^2 of 76.8 %, 63.9 %, and 69.1 % for O_3, $PM_{2.5}$, and PM_{10}, respectively. We conducted univariate regression analyses to investigate whether long-term exposure to O_3, $PM_{2.5}$, PM_{10} and distance to regulatory monitors differs by socioeconomic indicators, the percentages of residents who were children (0–17 years) or age 65+ years in 154 neighborhoods. We also examined the number of days exceeding the Brazilian National Air Quality Standard (BNAQS). Long-term exposures to O_3 and $PM_{2.5}$ were higher in more socially deprived neighborhoods. An interquartile range (IQR) increment of the social development index (SDI) was associated

*Corresponding author. 1086 SPHPI, MC 923, 1603 West Taylor St. Chicago, IL, 60612, United States. honghyok@uic.edu (H. Kim).
☆The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Declaration of competing interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement
Honghyok Kim: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ji-Young Son:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Washington Junger:** Writing – review & editing, Validation, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Michelle L. Bell:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2024.120374.

with a 3.6 μg/m^3 (95 % confidence interval [CI]: 2.9, 4.4; p-value≤0.001) decrease in O_3, and 0.3 μg/m^3 (95 % CI: 0.2, 0.5; p-value = 0.010) decrease in $PM_{2.5}$. An IQR increase in the percentage of residents who are children was associated with a 4.1 μg/m^3 (95 % CI: 3.1, 5.0; p-value≤0.001) increase in O_3, and 0.4 μg/m^3 (95 % CI: 0.3, 0.6; p-value = 0.009) increase in $PM_{2.5}$. An IQR increase in the percentage of residents age ≥65was associated with a 3.3 μg/m^3 (95 % CI: 2.4, 4.3; p-value=<0.001) decrease in O_3, and 0.3 μg/m^3 (95 % CI: 0.1, 0.5; p-value = 0.058) decrease in $PM_{2.5}$. There were no apparent associations for PM_{10}. The association for daily O_3 levels exceeding the BNAQS daily standard was 0.4 %p–0.8 %p different by the IQR of variables, indicating a 7–15 days difference in the six-year period. The association for daily $PM_{2.5}$ levels exceeding the BNAQS daily standard showed a 0.7–1.5 %p difference by the IQR, meaning a 13–27 days difference in the period. We did not find statistically significant associations between the distance to monitors and neighborhood characteristics but some indication regarding SDI. We found that O_3 levels were higher in neighborhoods situated farther from monitoring stations, suggesting that elevated levels of air pollution may not be routinely measured. Exposure disparity patterns may vary by pollutants, suggesting a complex interplay between environmental and socioeconomic factors in environmental justice.

Keywords

Environmental justice; Air pollution modeling; Air pollution disparity; Information disparity; Socioeconomic inequality; Climate change

1. Introduction

Many studies have found that exposure to ambient air pollution may differ by sociodemographic factors. Globally, most of these studies focus on North America and Europe, demonstrating generally consistent results in North America but mixed evidence in Europe (Hajat et al., 2015). For example, a study on 215 census tracts in the United States for the years 2000–2006 demonstrated that particulate matter (PM) with aerodynamic diameter no larger than 2.5 μm ($PM_{2.5}$) concentration was higher with higher percentage of people without high school diploma, higher unemployment rate, and higher poverty rate (Bell and Ebisu, 2012). The concentration was higher for the non-Hispanic African American population and Hispanic population compared to the non-Hispanic White population. Similar results were observed in recent study on the continental United States for the years 2000–2016 (Jbaily et al., 2022). That study revealed that lower household income groups were exposed to higher levels of air pollution and that communities with higher-than-average percentage of White and Native American populations had lower air pollution than communities with higher-than-average percentage of residences who were Black, Asian and Hispanic or Latino.

Despite a wealth of research showing environmental justice issues for air pollution, research on such disparities in South America is very limited compared to that in other regions[(Gouveia et al., 2022)]. A recent review of environmental justice studies in Latin America showed that studies are concentrated in Brazil, Mexico, and Chile and that air pollution is more severe in more socially deprived areas (Gouveia et al., 2022). However, findings are mixed depending on spatial resolution of exposure and study populations. For example, a study on São Paulo, Brazil found that the highest socioeconomic group identified by the 2010 Census lives in the most polluted areas, which was based on traffic density for the year of 2007 in census tracts and surrounding areas (Habermann et al., 2014).

Another understudied area is disparity of information about air pollution. Studies in Brazil (Bravo and Bell, 2011) and the United States (Miranda et al., 2011) found that the location of monitors, often designed for regulatory purposes, is associated with sociodemographic factors, implying potential injustice regarding information about air pollution. Communities near monitors have access to better information about their air pollution exposure, whereas communities far from monitor would have more uncertainty in estimates of air pollution exposure. Monitors are commonly installed in urban areas with high population density for population exposure measurement and regulatory purposes. Some areas may be less represented in the monitoring network, such as areas with the lower population density such as suburban areas, while areas with high emission sources, such as industrial zones or traffic-heavy region may receive greater attention in monitoring campaigns. Certain sociodemographic subgroups may disproportionately inhabit these areas.

Therefore, this study aims to investigate disparities in air pollution exposure and information about air pollution by sociodemographic factors in the city of Rio de Janeiro, the second-most populous (6.211.233 residents in 2022 (PANORAMA, 2024)) in Brazil with highly dense urban areas. While interventions such as promotion of less polluting fuels have improved particulate air quality, severe air pollution episodes occur in this area (Gioda et al., 2016). For example, the O_3 level is particularly important due to climate change's impact on high temperature and solar radiation, which influence ozone formation (Geraldino et al., 2020). From July 2014 to July 2016, 1-h O_3 levels in one community exceeded the Brazilian Air Quality Standards of 160 μg/m^3 185 times (Geraldino et al., 2020). This echoes a grave concern of the impact of climate change on air pollution in South America where temperature has increased an average rate of 0.2 °C/decade (WMO, 2021). As the climate becomes hotter and drier, the ambient environment will become more favorable to high levels of air pollution such as ground-level O_3 and PM through more frequent and severe wildfires, secondary pollutant formations and stagnation (Skea et al., 2022).

2. Methods

2.1. Air pollution data

Hourly O_3, PM_{10} and $PM_{2.5}$ data were obtained from the monitoring stations maintained by the Municipal Secretariat for the Environment of Rio de Janeiro. We calculated daily 24-h PM_{10}, daily 24-h $PM_{2.5}$, and daily maximum 8-h O_3, following the Brazilian National Air Quality Standards (BNAQS) and the World Health Organization's air quality guidelines.

We used air pollution data in two different ways as described below, because our analyses are two-folds: 1) for estimating daily O_3, $PM_{2.5}$, and PM_{10} and identifying associations of sociodemographic factors with estimated O_3, $PM_{2.5}$, and PM_{10} levels; 2) for identifying associations of sociodemographic factors with the distance to monitoring stations for O_3 and PM_{10}.

2.1.1. For estimating daily O_3, $PM_{2.5}$, and PM_{10} and identifying associations of sociodemographic factors with estimated O_3, $PM_{2.5}$, and PM_{10} levels—

Not all monitoring stations had continuous daily measurements for PM_{10}, $PM_{2.5}$, and O_3 in the study period (2012–2017). To select data from monitoring stations, we identified monitoring stations for which we could compute at least one monthly average for PM_{10}, $PM_{2.5}$, and O_3 in the study period (2012–2017). Pollutant monthly averages were computed with ≥ three or four daily measurements (i.e., at least one measurement per week) for months with four or five weeks. Otherwise, we treated computed values as missing. We computed pollutant daily averages for days with at least 75 % of hourly values. We used a total of 11 stations for PM_{10} (19,363 observations), 20 stations for $PM_{2.5}$ (5344 observations), and 18 stations for O_3 (18,384 observations) distributed across the city of Rio de Janeiro. Table S1 shows descriptive statistics. Fig. S1 shows locations of stations.

2.1.2. For identifying associations of sociodemographic factors with the distance to monitoring stations for O_3 and PM_{10}—

To ensure the reliability of continuous measurements for regulatory purposes, we selected data from monitoring stations with less than 15 % missing data in the study period (2012–2017). We computed pollutant daily averages for days with at least 75 % of hourly values. We used a total of eight stations for PM_{10} and eight stations for O_3, distributed across the city of Rio de Janeiro. They are co-located (Fig. S1). We did not conduct analyses for $PM_{2.5}$ because there is only one monitoring station that meets our criteria. For $PM_{2.5}$, more than 85 % of data were missing for the other stations.

2.2. Estimation of PM_{10}, $PM_{2.5}$, and O_3 levels

We developed machine-learning based models to estimate daily levels of PM_{10}, $PM_{2.5}$, and O_3 for a 500m grid in the city of Rio de Janeiro. We estimated long-term exposure to PM_{10}, $PM_{2.5}$, and O_3 for the years 2012–2017 using machine learning-based models that leveraged more spatially granular data (See Section 2.2.2), and then calculated neighborhood-level averages. A neighborhood is defined as “Bairro”, which is an administrative subdivision of municipalities in Brazil. We selected a 500m grid based on the geographical size of Rio de Janeiro (approximately 1182.3 km^2) and the number of neighborhoods (i.e., 163) and to reduce computational time.

The following four subsections (2.2.1 to 2.2.4) describe our methods.

2.2.1. Data consolidation—

To develop models, we first consolidated different data from satellite remote sensing, global reanalyses for atmospheric and land variables and land use information and more. Details for data consolidation are provided in Supplementary Materials. Table S2 shows a list of datasets. Briefly, we linked estimates of daily levels of meteorological factors (e.g., dew point temperature, forecast albedo, solar radiation), aerosol optical depth (AOD), greenness, and land use polygons/points/line segments and elevation from a Digital Surface Model (DSM) to pollutant measurements using Geographical Information System (GIS) techniques. The utility of these data is based on their relevance to air pollutant levels, supported by literature in this field (Di et al., 2019; Huang et al., 2017; Li et al., 2020; Requia et al., 2020). We employed spatial and temporal interpolation techniques to fill in missing values. We performed spatial interpolation by calculating mean values for neighboring cells and replacing missing values accordingly. Similarly, for temporal interpolation, mean values of temporally adjacent cells (i.e., temporal lags) were calculated and used to replace missing values.

2.2.2. Machine learning—

We used eXtreme Gradient Boosting (XGBoost) to estimate daily pollutant levels. XGBoost is a well-established machine learning method, namely a classification algorithm, that is based on decision trees and can model the complex relationship between predictor variables and the dependent variable. Predictions from multiple trees are summarized iteratively; the errors arising from a weak decision tree are corrected by a new decision tree. To avoid overfitting, we selected hyperparameters used to construct trees: *Eta* = 0.1, *Subsample* = 0.5, and *Colsample_bytree* = 0.5. *Eta* controls the learning rate, ranging from 0 to 1, the lower *Eta* value means a model more robust to overfitting. *Subsample* controls subsample ratio of the training instance; the value 0.5 means that a model is fitted using randomly selected half of the data, which can prevent overfitting. *Colsample_bytree* controls subsample ratio of columns when constructing each tree. The lower value than 1 means that a model may be more robust to overfitting. We also used other hyperparameters with values similar to these, and the R^2 in the cross-validation (Section 2.2.3) were comparable.

2.2.3. Validation—

To test model performance, we conducted 10-fold cross-validation. In this cross-validation approach, the entire dataset is divided to 10 subsets. Nine subsets are used as a training set and the remaining subset, referred to as a test set. This process is iterated through 10 combinations, where each time, different subsets serve as a training set and a different subset serves as a test set (e.g., training set = 1st, …,9th subsets – test set = 10th subset; training set = 1st, …,8th, 10th subsets – test set = 9th subset). Estimated pollutant levels of the test sets from 10 iterations were compared to measured pollutant levels. Fig. S2 shows scatter plots for estimated and measured pollutant levels and overall R^2, spatial R^2, and temporal R^2. The validation process demonstrated that there is good agreement between estimated levels and measured levels: overall R^2 of 76.8 %, 63.9 %, and 69.1 % for O_3, $PM_{2.5}$, and PM_{10}.

2.2.4. Estimation—

With the trained models, we estimated daily levels of PM_{10}, $PM_{2.5}$ and O_3 at a 500m grid to reduce computational time. The more granular datasets were resampled at a 500m grid using bilinear interpolation. Subsequently, estimated daily levels at a 500m grid were further resampled at a 1 km grid using bilinear interpolation. Finally, we computed averages in the study period, using the estimated daily levels at a 1 km grid. 1 km-gridded daily estimates were compared to daily measurement at monitoring stations, supporting the utility of the 500m-gridded estimates.

2.3. Neighborhood-level characteristics

We obtained data for the Social Development Index (SDI) (Cavallieri and Lopes, 2008) for the year 2010, the number of residents for one-year interval age groups for the year 2010, and income per capita for the year 2010 (Brazilian Real, BRL per capita), and energy consumption per capita (kWh per capita) for the years 2012–2017 from the Mayor's Office and Instituto Pereira Passos. Percentage of residents who were children (aged 0–17 years) and that of residents who were seniors (aged 65+ years) were calculated. We previously applied the data for SDI (Silveira and Junger, 2018), which was calculated based on eight indicators measured from the 2010 Census; please see Silveira and Junger (2018) for detail. This is a normalized index ranging from 0 to 1. We multiplied this by 1000 to enhance visualization and to reduce the number of decimal places. Higher values indicate the higher socioeconomic development.

2.4. Statistical analysis

We conducted neighborhood-level analyses to identify PM_{10} and O_3 disparities by neighborhood characteristics. We fitted univariate regression models and presented results based on the difference in the pollutant level associated with an interquartile range (IQR) increase in a neighborhood-level characteristic. We also performed neighborhood-level analyses to investigate disparities in the number of days exceeding daily PM_{10}, $PM_{2.5}$, or O_3 level of the 2018 BNAQS by neighborhood characteristics. The daily PM_{10} level (24-h average), $PM_{2.5}$ level (24-h average) and O_3 (8-h maximum) level of the BNAQS is 50μg/ m^3, 25μg/m^3, and 100μg/m^3, respectively (Siciliano et al., 2020). The number of days exceeding the BNAQS were calculated based on the average of the daily pollutant estimates at 500m grids within neighborhoods. The number of the neighborhoods in the city of Rio de Janeiro is 163. The number used for final analyses was 154 due to missing values of neighborhood characteristics.

We also conducted analysis to identify whether neighborhoods close to monitoring stations are different from neighborhoods far to monitoring stations (Section 2.1.2). We categorized neighborhoods by the difference: within 2.5 km from monitoring stations (Group 1); between 2.5 km and 5 km from monitoring stations (Group 2); between 5 km and 10 km from monitoring stations (Group 3); and beyond 10 km from monitoring stations (Group 4). Distances were calculated based on geocoded locations of monitoring stations (Section 2.1.2) and the centroid of the geographical boundaries of neighborhoods.

3. Results

Table 1 presents descriptive statistics for long-term exposures to PM_{10}, $PM_{2.5}$, and O_3 and neighborhood characteristics by distance from air pollution monitoring stations. O_3 levels were higher at locations situated farther from monitoring stations. No such pattern was found for PM_{10} and $PM_{2.5}$. Sociodemographic factors were generally similar across proximity to monitoring stations, while SDI and income variables were lower in neighborhoods located ≥2.5 km and <10 km from monitoring stations (Group 2 and Group 3) than neighborhoods located <2.5 km from monitoring stations (Group 1) and neighborhoods located ≥10 km from monitoring stations (Group 4). Table 2 presents spatial correlation among variables. O_3, PM_{10}, and $PM_{2.5}$ were positively correlated with one another. O_3 and $PM_{2.5}$ were negatively correlated with SDI and income variables. SDI, income per capita, energy consumption per capita, and the percentage of residents who were aged 65+ years were highly positively correlated with each other while these were negatively correlated with the percentage of residents who were aged 0–17 years. O_3 and

$PM_{2.5}$ was negatively correlated with SDI, income per capita, and energy consumption per capita.

Fig. 1 presents spatial distributions of long-term pollution levels and the percent of days exceeding air pollution levels of the BNAQS for the years 2012–2017. O_3 levels ranged from 31.5μg/m^3 to 66.3μg/m^3 with a mean of 51.7μg/m^3. PM_{10} levels ranged from 31μg/m^3 to 67.3μg/m^3 with a mean of 38.8μg/m^3, indicating that the long-term average of PM_{10} levels in all neighborhoods exceeded the BNAQS annual mean standar (Di et al., 2019) of 20μg/m^3. $PM_{2.5}$ levels ranged from 11.0μg/m^3 to 18.2μg/m^3 with a mean of 14.4μg/m^3. The long-term average of $PM_{2.5}$ levels in all neighborhoods also exceeded the standard, given that the BNAQS annual mean standard (Di et al., 2019) for $PM_{2.5}$ is 10μg/m^3. The average number of days with daily O_3 levels exceeding the BNAQS is 47 across the neighborhoods (median: 30 days per neighborhood), which is 2.6 % of days for the period 2012–2017 (a total of 1827 days), while for PM_{10} it was 375.8 (median: 323 days per neighborhood), which is 20.5 % of days for the period 2012–2017. For $PM_{2.5}$, the number was 157 (median: 147 days per neighborhood), which is 8.6 % of days for the period 2012–2017. The range of days exceeding the BNAQS across neighborhoods ranged from 0 days to 194 days (10.6 % of the days) for O_3; 108 days (5.9 % of the days) to 1790 days (98 % of the days) for PM_{10}; and 17 days (0.9 % of the days) to 314 days (17.2 % of the days) for $PM_{2.5}$.

Fig. 2 shows scatter plots for long-term exposure to O_3 and neighborhood characteristics. Linear regression coefficients, 95 % confidence intervals, and p-values are also presented in the plots. Higher SDI was associated with the lower O_3 level. Specifically, an IQR increment of SDI (71.8) was associated with a 3.6μg/m^3 (95 % confidence interval [CI]: 2.9, 4.4; p-value=<0.001) decrease in O_3. Similarly, the higher income per capita was associated with lower O_3. An IQR increment in the logarithm of income (BRL) per capita (0.6) was associated with a 3.7μg/m^3 (95 % CI: 3.1, 4.4; p-value=<0.001) decrease in O_3. The higher energy consumption per capita was further associated with lower O_3. An IQR increment in energy consumption per capita (0.9 kWh per capita) was associated with a 1.8μg/m^3 (95 % CI: 1.0, 2.5; p-value = 0.014) decrease in O_3. A higher percentage of residents who were aged 0–17 years was associated with higher O_3. An IQR increase (6.8 %p) was associated with a 4.1μg/m^3 (95 % CI: 3.1, 5.0; p-value=<0.001) increase in O_3. In contrast, an IQR increase in the percentage of residents who were aged 65+ (5.3 %p) was associated with a 3.3μg/m^3 (95 % CI: 2.4, 4.3; p-value=<0.001) decrease in O_3.

Fig. 3 shows scatter plots for long-term exposure to $PM_{2.5}$ and neighborhood characteristics. In general, patterns for $PM_{2.5}$ were consistent with those for O_3. Higher SDI was associated with the lower $PM_{2.5}$ level. Specifically, an IQR increment of SDI (71.8) was associated with a 0.3μg/m^3 (95 % confidence interval [CI]: 0.2, 0.5; p-value = 0.010) decrease in $PM_{2.5}$. Similarly, the higher income per capita was associated with lower $PM_{2.5}$. An IQR increment in the logarithm of income (BRL) per capita (0.6) was associated with a 0.5μg/m^3 (95 % CI: 0.4, 0.7; p-value=<0.001) decrease in $PM_{2.5}$. The higher energy consumption per capita was further associated with lower $PM_{2.5}$. An IQR increment in energy consumption per capita (0.9 kWh per capita) was associated with a 0.4μg/m^3 (95 % CI: 0.2, 0.5; p-value = 0.003) decrease in O_3. A higher percentage of residents who were aged 0–17 years was associated with higher $PM_{2.5}$. An IQR increase (6.8 %p) was associated with a 0.4μg/m^3

(95 % CI: 0.3, 0.6; p-value = 0.009) increase in $PM_{2.5}$. In contrast, an IQR increase in the percentage of residents who were aged 65+ (5.3 %p) was associated with a 0.3μg/m^3 (95 % CI: 0.1, 0.5; p-value = 0.058) decrease in $PM_{2.5}$.

Associations for long-term exposure to PM_{10} were not consistent with those for O_3 and $PM_{2.5}$. We did not find significant linear associations between neighborhood characteristics and PM_{10}; No clear patterns were shown (Fig. 4).

Figs. S3–5 shows scatter plots for the percent of days exceeding daily O_3 level, or daily $PM_{2.5}$ level, daily PM_{10} level of the BNAQS and neighborhood characteristics. Linear regression coefficients, 95 % confidence intervals, and p-values are also presented in the plots. The higher SDI was associated with fewer days exceeding daily O_3 level of the BNAQS. An IQR increment of SDI was associated with a 0.6 %p (95 % CI: 0.4, 0.8; p-value = 0.006) decrease in the percent of the days (11 days (95 % CI: 7, 15)). An IQR increment in the logarithm of income per capita was associated with a 0.8 %p (95 % CI: 0.6, 1.0; p-value=<0.001) decrease in the percent of the days (15 days (95 % CI: 11, 18)). The IQR increment in energy consumption per capita was associated with a 0.4 % p (95 % CI: 0.2, 0.6; p-value = 0.036) decrease in the percent of the days (7 days (95 % CI: 4, 11)). An IQR increase in the percentage of residents who were children was associated with a 0.7 %p (95 % CI: 0.4, 1.0; p-value = 0.014) increase in the percent of the days (13 days (95 % CI: 7, 18). In contrast, the IQR increase in the percentage of residents age 65+ years was associated with a 0.4 %p (95 % CI: 0.1, 0.7; p-value = 0.139) decrease in the percent of days (7 days (95 % CI: 2, 13)). Consistent patterns were identified for $PM_{2.5}$. We found that a 0.7–1.5 %p difference in the percent of days exceeding the BNAQS daily $PM_{2.5}$ standard by the IQR of the variables, meaning a 13–27 days difference in the period. No apparent patterns were found for PM_{10}.

Table 3 presents whether neighborhood characteristics differed based on their proximity to a monitoring station, with distance between neighborhood centroid and monitor divided into four groups. We did not find statistically significant associations between the distance to monitor and neighborhood characteristics, except that the percentage of residents aged 65+ years for Group 3 (5–10 km) differed by −1.8 %p (95 % CI: 3.4, −0.2) compared to that of Group 1. Although SDI was not statistically associated with the distance, SDI of Group 1 (≤2.5 km) was higher than that of Group 2 (−24.2 [95 % CI: 51.8, 3.4; p-value = 0.088]) and Group 3 (−19.5 [95 % CI: 45.8, 6.8; p-value = 0.148]). Although results were not statistically significant, general trends were that with further distance from monitors having lower SDI and fewer seniors.

4. Discussion

The findings of our study indicate that disparities in air pollution may vary depending on the specific pollutants being analyzed. Our findings rely on our novel, validated gridded estimates of daily levels of $PM_{2.5}$, PM_{10}, and O_3. We leveraged machine learning algorithms and diverse datasets from satellite remote sensing, global reanalyses for atmospheric and land variables to estimate pollutant levels. Factors associated with high socioeconomic status, such as SDI, income per capita, and energy consumption per capita, showed negative association with long-term exposure to $PM_{2.5}$ and O_3. The association for a six-year average of O_3 levels was approximately 1.8–4.1μg/m^3 different by an IQR in these variables whereas that for daily O_3 levels exceeding the BNAQS daily standard was 0.4 %p–0.8 %p different by the IQR, indicating a 7–15 days difference in the six-year period. This difference may be seen as substantial as the daily standard for O_3 was, on average, exceeded on 2.6 % of the days (i.e., 30 days per neighborhood) in the city of Rio de Janeiro. The association for a six-year average of $PM_{2.5}$ levels differed by 0.3–0.4μg/m^3 for the IQR. The face value may not appear significant, but its importance becomes evident when considering potential health effects. For instance, this disparity could result in a 0.18–0.64 % difference in excess mortality risk between neighborhoods when individuals are exposed to $PM_{2.5}$ for ≥10 years. This is calculated by multiplying the 6–16 % increase in mortality risk per 10μg/m^3 due to a decade-long exposure to $PM_{2.5}$ (refer to Fig. 2 in Brook et al. (2010)) by 0.03 or 0.04, resulting in a range of 0.18–0.64 %. This may further translate to the disparity in substantial excess deaths attributable to $PM_{2.5}$ given that the city of Rio de Janeiro is the second most populous city in Brazil. The association for daily $PM_{2.5}$ levels exceeding the BNAQS daily standard showed a 0.7–1.5 %p difference by the IQR, meaning a 13–27 days difference in the period. While PM_{10} were positively correlated with O_3 and $PM_{2.5}$, we did not find apparent patterns of PM_{10} regarding neighborhood characteristics. Our findings are consistent with some nationwide studies for the United States (Bell and Ebisu, 2012; Jbaily et al., 2022), but are not consistent with some studies for Sao Paulo, Brazil, (Habermann et al., 2014) and individual cities of European countries (Hajat et al., 2015) and the United States (Huang et al., 2019; Shmool et al., 2016). The correlation between PM levels and O_3 levels varies by location. Observations may relate to the demographic patterns of the highly urbanized environment.

We found that all neighborhoods experienced exceedance of the BNAQS annual standard for PM_{10} and $PM_{2.5}$ levels. On average, the BNAQS daily mean standard for $PM_{2.5}$ was exceeded on 8.5 % of the days for the period 2012–2017 and for PM_{10}, it was exceeded on 20 % of the days during the same period. Alarmingly, one neighborhood’s daily PM_{10} levels exceeded the standard on 97 % of the days. There is no annual standard for O_3 in the BNAQS, but on average, the daily 8-h mean standard for O_3 was exceeded on 2.6 % of the days. However, the number of days exceeding the daily standards varies significantly across neighborhoods. Similar to the long-term averages of $PM_{2.5}$ and O_3, the occurrences of days exceeding the daily standards for $PM_{2.5}$ and O_3 were associated with sociodemographic factors.

In Rio de Janeiro, the presence of heavy traffic contributes to PM and nitrogen oxides (NOx) emissions leading to air pollution. Many individuals use their vehicles for transportation. It is likely that individuals in neighborhoods with high socioeconomic status may own and use vehicles more frequently, compared to those in neighborhoods with low socioeconomic status. Risk communication and risk governance may merit attention to reduce auto vehicle emissions and to promote the improvement of public transportation systems. Other possibilities of high levels of air pollution in local neighborhoods include proximity to industrial facilities.

We found that neighborhoods with the higher percentage of residents who were children are exposed to higher long-term exposure to O_3 and experienced more days exceeding daily O_3 BNAQS. These results mean that a large number of children is exposed to both short-term and long-term exposure to O_3. Short-term exposure to O_3 causes exacerbation of asthma and chronic obstructive pulmonary diseases (COPD) in children and long-term exposure to O_3 may lead to development of asthma and COPD (USEPA, 2020). Short-term exposure to O_3 may be related to compromised lung function and growth in children (Shmool et al., 2016).

Our results suggest that all neighborhoods might have experienced at least one exceedance of the BNAQS daily standards during the six-year period although the pollution levels were estimated. This is concerning, given that short-term exposure to PM and O_3 is associated with an increased risk of various health outcomes, including mortality (Orellano et al., 2020), hospitalization (Bell et al., 2013, 2014; Gao et al., 2020; Shah et al., 2015; Yee et al., 2021; Zheng et al., 2021), and emergency department visits (Bell et al., 2014; Yee et al., 2021; Zheng et al., 2021). Particularly alarming are the disparities in $PM_{2.5}$ and O_3 exposures–the higher levels for communities with lower socioeconomic status, because the population with the lower socioeconomic status may be the most vulnerable subgroup due to the higher exposure and possibly higher increase in health risks per a given level of exposure (e.g., effect modification in epidemiology) (Bell et al., 2014). We also identified the long-term averages of $PM_{2.5}$ and PM_{10} exceeded the BNAQS annual standards in all neighborhoods.

Although we did not find statistically significant findings in relation to socioeconomic characteristics and proximity to air pollution monitoring stations, we observed some general trends. In this city, monitors were more likely to be located near neighborhoods with higher SDI (≤2.5 km vs. 2–5 km and 5–10 km, Table 3). This means that the air pollution exposure levels of some neighborhoods may be less represented by the measurements from monitoring stations compared to others, although for this city the least susceptible neighborhoods are better represented. Furthermore, we discovered that O_3 levels were negatively associated with the proximity to monitoring stations; higher O_3 levels in neighborhoods situated farther from stations. This trend may be concerning because pollution levels in such neighborhoods may not be routinely measured. Estimates of exposure to air pollution in health studies are often calculated based on the measurements from monitoring stations, such that different proximity to monitors by subpopulation could introduce exposure error that differs by subpopulation (Bravo and Bell, 2011). Such exposure errors could be critical when the study is designed to investigate different effects across subpopulations. Thus, a broader representation of air pollution levels across neighborhoods is important for public health protection as well as air pollution and health studies.

We acknowledge several limitations of this study. Firstly, while we developed machine learning based models that can estimate local levels of air pollutants, there exist still uncertainties in daily level estimates. Overall R^2 was 76.8 %, 63.9 %, and 69.1 % for O_3, $PM_{2.5}$, and PM_{10}, respectively and uncertainties may have contributed to our findings. Such uncertainties originate from multiple sources such as input datasets, a coverage of monitoring campaigns, and computational complexity. While datasets from multiple sources, including satellite remote sensing and global reanalysis/assimilation for atmospheric conditions, can enhance model performance, such data typically have high spatial/temporal resolution but low temporal/spatial resolution. For example, two different datasets for AOD estimates complemented each other: Modern-Era Retrospective analysis for Research and Applications Version 2 (approximately 0.55° every 3 h); Moderate Resolution Imaging Spectroradiometer Multi-Angle Implementation of Atmospheric Correction from Terra and Aqua Satellites (1 km; monthly and yearly values). As shown in this study, while utilizing various datasets within the framework of machine learning algorithms can enhance the benefits of such data, incorporating even more spatially and temporally granular information would further improve the accuracy of air pollutant level estimates. Moreover, more expansive monitoring campaigns would provide higher spatial coverage of and less-error prone air pollution estimates and potentially more accurate for locations far from the current monitors, which could aid evidence regarding exposure disparities. Second, the SDI and demographic variables were calculated based on the 2010 Census, which differs from the years of air pollution data (2012–2017). Our analysis assumes that changes in sociodemographic characteristics over time are minimal. We were unable to consider more recent air pollution and sociodemographic data due to data availability issues. Third, we considered only PM_{10}, $PM_{2.5}$, and O_3, not nitrogen oxides, carbon monoxide, and other hazardous air pollutants (e.g., chemical compositions, volatile organic compounds). Fourth, race/ethnicity was not considered. We focused on socioeconomic inequalities and age structure inequalities, while racial/ethnic composition of neighborhood populations may be highly diverse and may be related to socioeconomic factors. In Rio de Janeiro, there are many favelas where most of the poor Black population live but sharing the same neighborhood of very rich people. Measuring segregation with hyper local estimates of race/ethnicity may be more desirable than race/ethnicity indicator alone. Future research is needed on these topics.

In conclusion, our study demonstrates that long-term exposure to O_3 and $PM_{2.5}$ differed by sociodemographic factors in the city of Rio de Janeiro for the years 2012–2017. The estimated daily levels of PM_{10}, $PM_{2.5}$ and O_3 often exceeded the BNAQS and long-term averages of PM_{10} and $PM_{2.5}$ consistently surpassed the BNAQS annual standard in all neighborhoods, which raises concerns regarding the health effects of short-term and long-term exposures to high levels of these pollutants. The number of days exceeding the BNAQS daily standard for O_3 and $PM_{2.5}$ also varied based on sociodemographic factors. This study addresses air pollution and sociodemographic disparities in an underrepresented location, with findings that differed from that of some other cities. This indicates the need for local studies in many areas to investigate patterns of environmental justice in relation to air pollution exposure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This publication was developed under Assistance Agreement No. RD835871 awarded by the U.S. EPA to Yale University. It has not been formally reviewed by the U.S. EPA. Research reported in this publication was also supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health under Award Number R01MD012769.

This work was supported by the Wellcome Trust [216087/Z/19/Z]. Dr Honghyok Kim was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A3A14039711).

Data availability

Data will be made available on request.

References

Bell ML, Ebisu K, 2012. Environmental inequality in exposures to airborne particulate matter components in the United States. Environ. Health Perspect. 120 (12), 1699–1704. [PubMed: 22889745]

Bell ML, Zanobetti A, Dominici F, 2013. Evidence on vulnerability and susceptibility to health risks associated with short-term exposure to particulate matter: a systematic review and meta-analysis. Am. J. Epidemiol. 178 (6), 865–876. [PubMed: 23887042]

Bell ML, Zanobetti A, Dominici F, 2014. Who is more affected by ozone pollution? A systematic review and meta-analysis. Am. J. Epidemiol. 180 (1), 15–28. [PubMed: 24872350]

Bravo MA, Bell ML, 2011. Spatial heterogeneity of PM10 and O3 in São Paulo, Brazil, and implications for human health studies. J. Air Waste Manag. Assoc. 61 (1), 69–77. [PubMed: 21305890]

Brook RD, Rajagopalan S, III Pope C.A., et al. , 2010. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. Circulation 121 (21), 2331–2378. [PubMed: 20458016]

Cavallieri F, Lopes GP, 2008. Índice de Desenvolvimento Social-IDS: comparando as realidades microurbanas da cidade do Rio de Janeiro. Coleção Estudos Cariocas 8 (20080401), 12.

Di Q, Amini H, Shi L, et al. , 2019. An ensemble-based model of PM2. 5 concentration across the contiguous United States with high spatiotemporal resolution. Environ. Int. 130, 104909. [PubMed: 31272018]

Gao H, Wang K, Au W W, Zhao W, Xia Z-l, 2020. A systematic review and meta-analysis of short-term ambient ozone exposure and COPD hospitalizations. Int. J. Environ. Res. Publ. Health 17 (6), 2130.

Geraldino CGP, Arbilla G, da Silva CM, Corrêa SM, Martins EM, 2020. Understanding high tropospheric ozone episodes in Bangu, Rio de Janeiro, Brazil. Environ. Monit. Assess. 192 (3), 156. [PubMed: 32008106]

Gioda A, Ventura LMB, Ramos MB, Silva MPR, 2016. Half century monitoring air pollution in a Megacity: a case study of Rio de Janeiro. Water, Air, Soil Pollut. 227, 1–17.

Gouveia N, Slovic AD, Kanai CM, Soriano L, 2022. Air Pollution and environmental justice in Latin America: where are we and how can we move forward? Current Environmental Health Reports 9 (2), 152–164. [PubMed: 35146705]

Habermann M, Souza M, Prado R, Gouveia N, 2014. Socioeconomic inequalities and exposure to traffic-related air pollution in the city of São Paulo, Brazil. Cad. Saúde Pública 30, 119–125. [PubMed: 24627019]

Hajat A, Hsia C, O'Neill MS, 2015. Socioeconomic disparities and air pollution exposure: a global review. Current Environmental Health Reports 2, 440–450. [PubMed: 26381684]

Huang L, Zhang C, Bi J, 2017. Development of land use regression models for PM2. 5, SO2, NO2 and O3 in Nanjing, China. Environ. Res. 158, 542–552. [PubMed: 28715783]

Huang K, Bi J, Meng X, et al. , 2019. Estimating daily PM2. 5 concentrations in New York City at the neighborhood-scale: implications for integrating non-regulatory measurements. Sci. Total Environ. 697, 134094. [PubMed: 32380602]

Jbaily A, Zhou X, Liu J, et al. , 2022. Air pollution exposure disparities across US population and income groups. Nature 601 (7892), 228–233. [PubMed: 35022594]

Li L, Girguis M, Lurmann F, et al. , 2020. Ensemble-based deep learning for estimating PM2. 5 over California with multisource big data including wildfire smoke. Environ. Int. 145, 106143. [PubMed: 32980736]

Miranda ML, Edwards SE, Keating MH, Paul CJ, 2011. Making the environmental justice grade: the relative burden of air pollution exposure in the United States. Int. J. Environ. Res. Publ. Health 8 (6), 1755–1771.

Orellano P, Reynoso J, Quaranta N, Bardach A, Ciapponi A, 2020. Short-term exposure to particulate matter (PM10 and PM2. 5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: systematic review and meta-analysis. Environ. Int. 142, 105876. [PubMed: 32590284]

PANORAMA, 2024. Population in the City of Rio de Janeiro. https://censo2022.ibge.gov.br/panorama. (accessed 3 February 2024).

Requia WJ, Di Q, Silvern R, et al. , 2020. An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. Environ. Sci. Technol. 54 (18), 11037–11047. [PubMed: 32808786]

Shah AS, Lee KK, McAllister DA, et al. , 2015. Short term exposure to air pollution and stroke: systematic review and meta-analysis. BMJ 350.

Shmool JLC, Kinnee E, Sheffield PE, Clougherty JE, 2016. Spatio-temporal ozone variation in a case-crossover analysis of childhood asthma hospital visits in New York City. Environ. Res. 147, 108–114. [PubMed: 26855129]

Siciliano B, Dantas G, Silva CMD, Arbilla G, 2020. The updated Brazilian national air quality standards: a critical review. J. Braz. Chem. Soc. 31, 523–535.

Silveira IHd, Junger WL, 2018. Green spaces and mortality due to cardiovascular diseases in the city of Rio de Janeiro. Rev. Saude Publica 52.

Skea J, Shukla P, Kılkış Ş, 2022. Climate Change 2022: Mitigation of Climate Change. Cambridge University Press, Cambridge (MA), USA.

United States Environmental Protection Agency, 2020. Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants.

World Meteorological Organization, 2021. State of the Climate in Latin America and the Caribbean.

Yee J, Cho YA, Yoo HJ, Yun H, Gwak HS, 2021. Short-term exposure to air pollution and hospital admission for pneumonia: a systematic review and meta-analysis. Environ. Health 20 (1), 1–10. [PubMed: 33407552]

Zheng X-y, Orellano P, Lin H-l, Jiang M, Guan W-j, 2021. Short-term exposure to ozone, nitrogen dioxide, and sulphur dioxide and emergency department visits and hospital admissions due to asthma: a systematic review and meta-analysis. Environ. Int. 150, 106435. [PubMed: 33601224]

HIGHLIGHTS

- We developed hyper local estimates of daily PM_{10}, $PM_{2.5}$ and O_3 levels.
- $PM_{2.5}$ and O_3 levels were higher in more socially deprived neighborhoods.
- $PM_{2.5}$ and O_3 levels were higher in neighborhoods with more children.
- O_3 levels were higher in neighborhoods situated farther from monitoring stations.
- Levels exceeded air quality standards in many neighborhoods without stations.

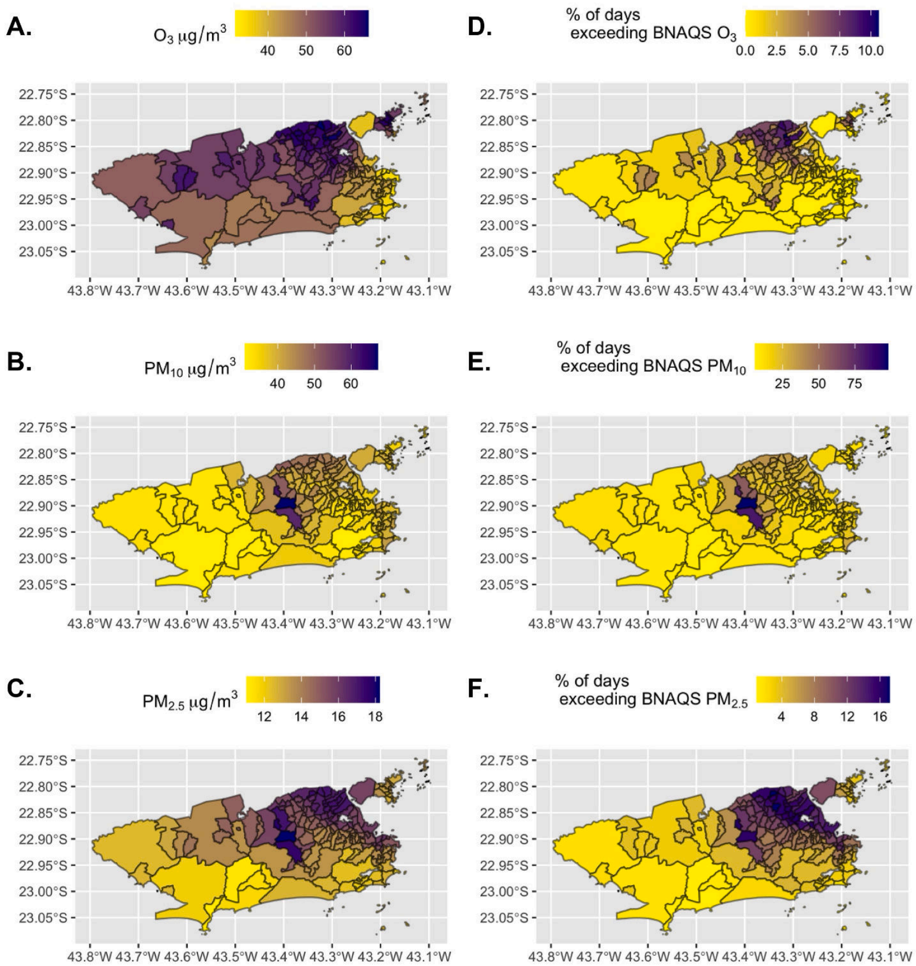

Fig. 1.

Spatial distribution of long-term exposure to air pollution and the percent of days exceeding air pollution levels of Brazilian National Ambient Quality Standard (BNAQS) in the city of Rio de Janeiro (2012–2017). Note: White shaded neighborhoods are omitted in the final analysis due to a missing value of at least one neighborhood variable. A total of 154 neighborhoods are included in the final analysis.

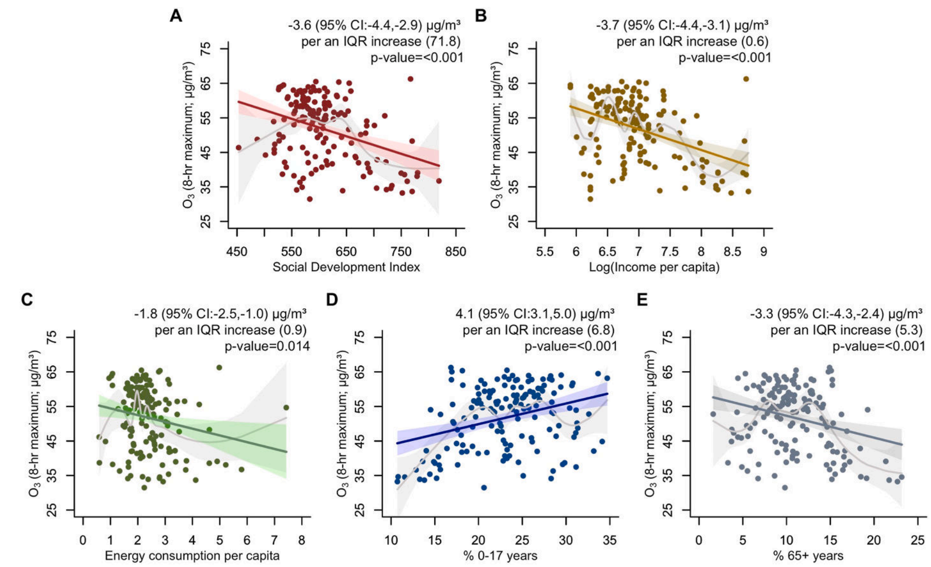

Fig. 2.

Scatter plots for neighborhood characteristics and long-term exposure to O_3 and for the years 2012–2017 in the city of Rio de Janeiro. Linear coefficients (95 % confidence intervals) per an IQR increase and their p-values are shown. Colored linear lines and adjacent shaded areas indicate predicted values and 95 % confidence intervals by linear coefficients. Grey curves and adjacent shaded areas indicate predicted values and 95 % confidence intervals by smoothing splines from generalized additive models (smoothing parameters were chosen by the generalized cross validation.

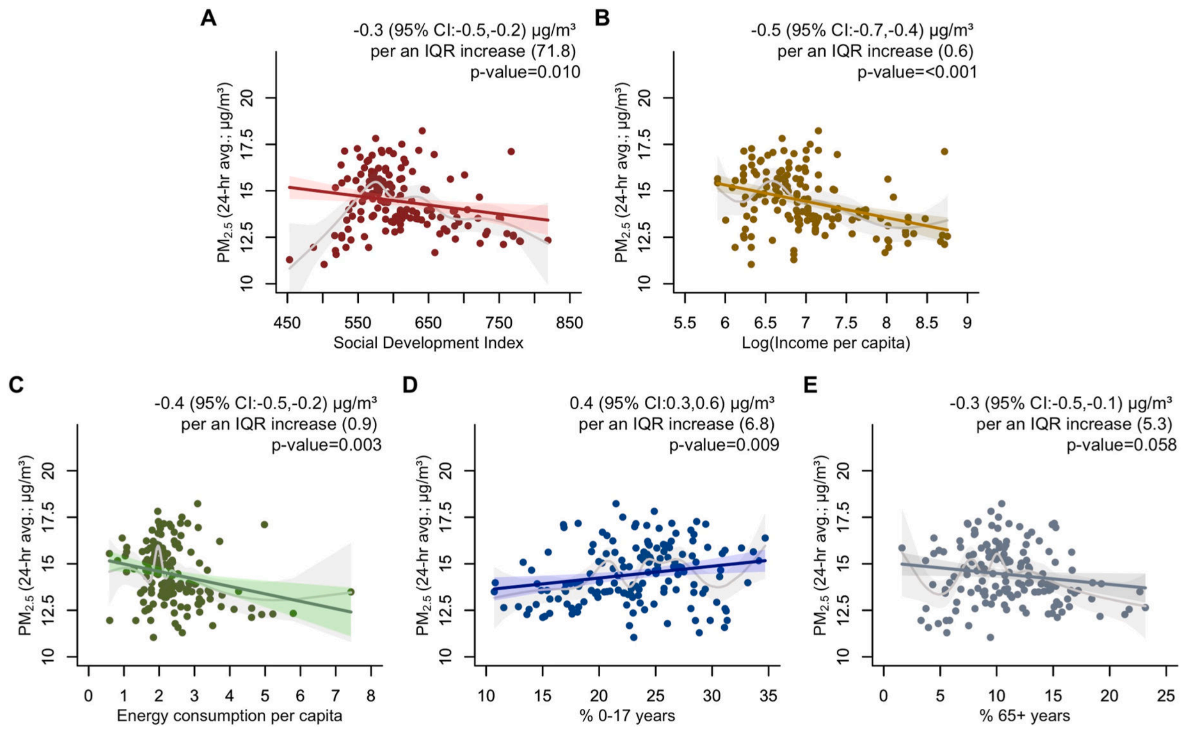

Fig. 3.

Scatter plots for neighborhood characteristics and long-term exposure to $PM_{2.5}$ and for the years 2012–2017 in the city of Rio de Janeiro. Linear coefficients (95 % confidence intervals) per an IQR increase and their p-values are shown. Colored linear lines and adjacent shaded areas indicate predicted values and 95 % confidence intervals by linear coefficients. Grey curves and adjacent shaded areas indicate predicted values and 95 % confidence intervals by smoothing splines from generalized additive models (smoothing parameters were chosen by the generalized cross validation.

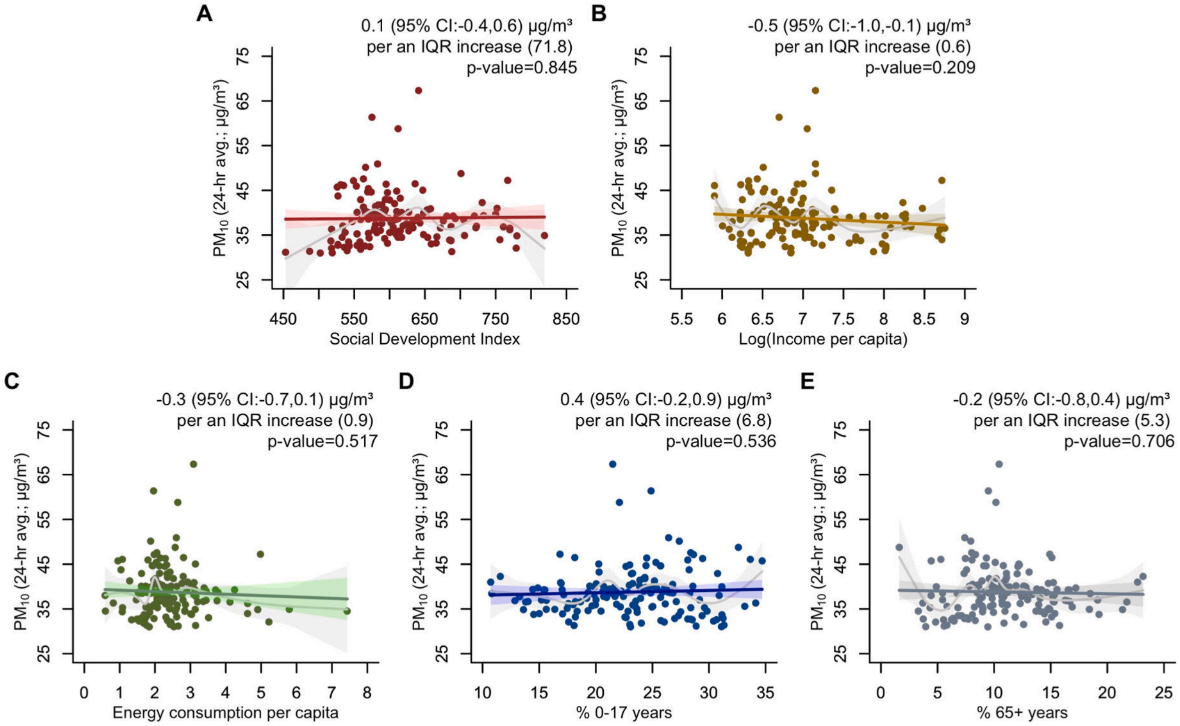

Fig. 4.

Scatter plots for neighborhood characteristics and long-term exposure to PM_{10} and for the years 2012–2017 in the city of Rio de Janeiro. Linear coefficients (95 % confidence intervals) per an IQR increase and their p-values are shown. Colored linear lines and adjacent shaded areas indicate predicted values and 95 % confidence intervals by linear coefficients. Grey curves and adjacent shaded areas indicate predicted values and 95 % confidence intervals by smoothing splines from generalized additive models (smoothing parameters were chosen by the generalized cross validation.

Table 1

Descriptive statistics of neighborhood characteristics and air pollutants (n = 154).

Mean (Standard deviation)									
Distance from air monitor	***n***	**O_3(μg/ m^3)**	**PM_{10} (μg/ m^3)**	**$PM_{2.5}$ (μg/ m^3)**	**Social Development Index**	**Income (BRL) per capita**	**Energy consumption (kwH) per capita**	**% aged 0–17 years**	**% aged 65+ years**
<2.5 km	53	46.7 (11.8)	38.4 (4.5)	14.7 (1.6)	620.0 (78.4)	1617 (1471)	2.4 (0.9)	22.1 (6.3)	11.9 (4.6)
2.5–5 km	38	47.7 (7.7)	38.6 (3.4)	14.9 (1.1)	595.8 (52.7)	1130 (1066)	2.2 (0.8)	24.0 (4.4)	10.6 (3.5)
5–10 km	45	54.3 (7.0)	39.6 (8.1)	14.1 (1.8)	600.4 (66.0)	1293 (960)	2.5 (1.1)	23.8 (4.7)	10.1 (3.8)
>10 km	18	57.1 (5.8)	38.1 (2.5)	13.5 (0.7)	629.4 (50.9)	1558 (1205)	2.6 (0.8)	22.0 (3.9)	10.9 (3.3)

Table 2

Pearson correlation between neighborhood characteristics and air pollutants.

	PM_{10}	$PM_{2.5}$	Social Development Index	Income per capita	Energy consumption per capita	% aged 0–17 years	% aged 65+ years
O_3	0.34	0.41	−0.37	−0.43	−0.20	0.33	−0.28
PM_{10}		0.73	0.02	−0.10	−0.05	0.05	−0.03
$PM_{2.5}$			−0.21	−0.34	−0.24	0.21	−0.15
Social Development Index				0.78	0.74	−0.85	0.79
Income per capita					0.63	−0.64	0.62
Energy consumption per capita						−0.64	0.57
% aged 0–17 years							−0.92

Table 3

Difference in a neighborhood characteristic variable by distance from air pollution monitoring stations.

Neighborhood characteristics	Distance from air monitor	Increase (95 % confidence interval)	p-value
Social Development Index	≤2.5 km	Ref.	
	2.5–5 km	−24.2 (−51.8, 3.4)	0.088
	5–10 km	−19.5 (−45.8, 6.8)	0.148
	>10 km	9.5 (−25.9, 44.9)	0.601
Logarithm of income	≤2.5 km	Ref.	
per capita	2.5–5 km	−0.2 (−0.5, 0)	0.069
	5–10 km	− 0.1 (−0.3, 0.2)	0.637
	>10 km	0.1 (−0.2, 0.5)	0.457
Energy consumption	≤2.5 km	Ref.	
per capita	2.5–5 km	− 0.1 (−0.5, 0.3)	0.516
	5–10 km	0.2 (−0.2, 0.5)	0.415
	>10 km	0.2 (−0.3, 0.7)	0.389
% aged 0–17 years	≤2.5 km	Ref.	
	2.5–5 km	1.8 (−0.3, 4.0)	0.097
	5–10 km	1.6 (−0.4, 3.7)	0.120
	>10 km	−0.1 (−2.9, 2.6)	0.917
% aged 65+ years	≤2.5 km	Ref.	
	2.5–5 km	−1.2 (−2.9, 0.4)	0.146
	5–10 km	−1.8 (−3.4, −0.2)	0.030
	>10 km	−0.9 (−3.1, 1.2)	0.393