



## **Senseable City Lab :::: Massachusetts Institute of Technology**

This paper might be a pre-copy-editing or a post-print author-produced .pdf of an article accepted for publication. For the definitive publisher-authenticated version, please refer directly to publishing house's archive system

# Big mobility data reveals hyperlocal air pollution exposure disparities in the Bronx, New York

Received: 11 November 2023

Accepted: 18 June 2024

Published online: 29 July 2024

 Check for updates

Iacopo Testi<sup>1,8</sup>, An Wang<sup>1,2,8</sup>✉, Sanjana Paul<sup>1</sup>, Simone Mora<sup>1,3</sup>✉, Erica Walker<sup>4</sup>, Marguerite Nyhan<sup>5</sup>, Fábio Duarte<sup>1</sup>, Paolo Santi<sup>1,6</sup> & Carlo Ratti<sup>1,7</sup>

Air pollution disproportionately affects socially disadvantaged populations. Our study bridges the existing gap in quantifying mobility-based exposure and its associated disparity issues. We combined the granular mobility of over 500,000 unique anonymized users daily and hyperlocal air pollution data in 100 × 100-m grid cells to quantify disparities in particulate matter exposure in a racially diverse and dense urban area of New York City. Our approach advances the study of exposure and its disparity from individualized exposure tracking to a population-representative scale. We observed apparently different spatial patterns between personal exposure and exposure disparities, noting that people from Hispanic-majority and low-income neighborhoods were those most severely and disproportionately exposed to fine particulate matter (PM<sub>2.5</sub>) pollution. We reveal that race and ethnicity are much stronger indicators of exposure disparity than income. Our study further demonstrates that within-group variation contributes a major portion to exposure disparities, suggesting more granular mitigation plans are needed to target high-exposure individuals from socially disadvantaged groups in addition to generic air quality improvement.

Ethnic minorities and economically disadvantaged communities in the United States are disproportionately exposed to air pollution<sup>1</sup>. Despite an overall improvement in air quality throughout the country<sup>2,3</sup>, ethnic and economic disparities persist, affecting health outcomes and the quality of life of disadvantaged groups<sup>4–7</sup>. Quantifying disparities in air pollution exposure is crucial as a means to address cumulative impacts, define guiding principles for mitigation, facilitate resource investment to promote health in overexposed communities, improve community-level environmental health literacy and awareness, and provide data to direct and inform equitable policy interventions.

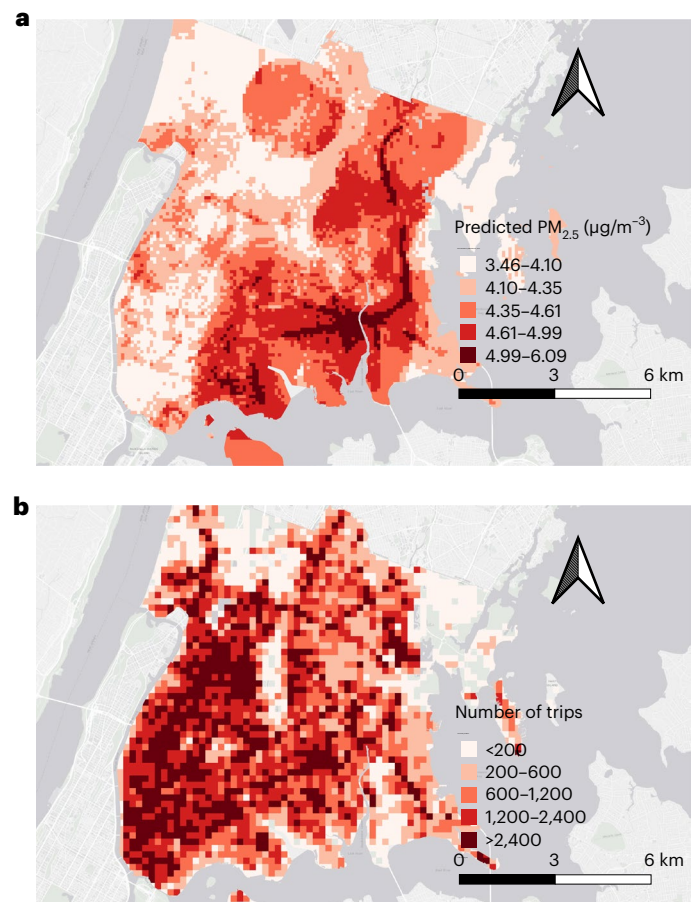
A recent upsurge in information in the scientific literature has revealed more profound deep-rooted disparities in air pollution exposure among vulnerable groups. The population-weighted fine particulate matter (PM<sub>2.5</sub>) exposure per emissions source among communities has been estimated<sup>8</sup>, showing that nearly all major emissions categories are disproportionately and systemically affecting people of color. Brazil<sup>9</sup> found that socially disadvantaged populations' daily mobility patterns exacerbated their exposure to air pollution. A more concerning fact is that, while air pollution levels continue to decrease across the United States, relative exposure disparities among racial and ethnic

<sup>1</sup>Senseable City Lab, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>2</sup>Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China. <sup>3</sup>Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway. <sup>4</sup>School of Public Health, Brown University, Providence, RI, USA. <sup>5</sup>School of Engineering and Architecture, University College Cork, Cork, Ireland. <sup>6</sup>Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche, Pisa, Italy. <sup>7</sup>ABC Department, Politecnico di Milano, Milan, Italy. <sup>8</sup>These authors contributed equally: Iacopo Testi, An Wang. ✉e-mail: [an.wang@polyu.edu.hk](mailto:an.wang@polyu.edu.hk); [oras@mit.edu](mailto:oras@mit.edu)

groups are growing<sup>1</sup>. It has been discovered that historical redlining—a discriminatory mortgage appraisal practice from the 1930s—has been a long-lasting contributor to systemic air pollution exposure disparities in the United States<sup>10</sup>. These disparities have occurred because there is substantially less green space in redlined areas, sources of air pollution being situated near redlined communities, and there is also greater exposure to vehicular traffic emissions<sup>11,12</sup>. A growing family of environmental justice scoring and mapping tools, created by US governmental agencies, is available at the state and national levels. Most of these intersect static environmental burdens and population characteristics to define synthesized environmental justice scores, including the CalEnviroScreen tool ([oehha.ca.gov/calenviroscreen/](http://oehha.ca.gov/calenviroscreen/)), the EJScreen ([epa.gov/ejscreen](http://epa.gov/ejscreen)) and the EJ Index tool ([atsdr.cdc.gov/placeandhealth/eji](http://atsdr.cdc.gov/placeandhealth/eji)). However, a caveat has emerged from the existing literature and these tools concerning the exposure disparities being explored using static, population-weighted, residence-based methods. In the absence of information on individuals' mobility patterns, the places that people visit have been overlooked, which leads to two latent questions: (1) Does the inclusion of high-resolution mobility data enhance the accuracy of exposure estimation? (2) Does it matter where the exposure occurs or is it just the total amount of exposure that matters?

There is no doubt that there is huge variability in people's mobility patterns, especially in the amount of time that people spend away from home<sup>13,14</sup>. However, there seems to be a lack of consensus in previous studies on whether the variability in people's mobility trickles downstream to exposure estimation to an extent that it is worth the effort of including such mobility data. Kwan<sup>15</sup> was the first to highlight the neighborhood effect averaging problem, stating that the residence-based estimation of environmental factor exposure (for example, air pollution and noise) can generate more polarized results than a mobility-based estimation. However, Desouza et al.<sup>16</sup> and Reis et al.<sup>17</sup> estimated the PM<sub>2.5</sub> exposure difference between residence-only and workplace-included methods in the United States and United Kingdom, respectively, both observing very small differences in population-weighted PM<sub>2.5</sub> concentrations between the two methods (0.6% in the United States and 0.3% in the United Kingdom), except for a very few corner cases. Similar results have also been observed in more localized studies in New York City<sup>18,19</sup>, where mobile device traffic data at the cell-tower level has been used. It has been reported that traffic-related air pollution with short transport distances, such as nitrogen dioxide (NO<sub>2</sub>), can lead to more-manifest differences between residence- and mobility-based methods<sup>17,20</sup>. From an epidemiological perspective, the story is likely over because the total amount of exposure has stayed the same, leading to statistically similar health outcomes at an aggregated level. However, a recent study<sup>21</sup> demonstrated that emissions reduction at some (specific) locations is more effective than in other places in reducing national exposure disparities. Furthermore, there is an increasing body of literature indicating that city-level spatiotemporal variability in PM<sub>2.5</sub> compositions from diverse emissions sources<sup>22–24</sup> possess various toxic potencies at equal mass concentrations<sup>25–28</sup>. In this case, the question 'Where does exposure occur?' kicks in, denoting the utility of including mobility patterns in exposure studies for regulating emissions sources at strategic locations and accounting for health outcomes from certain PM<sub>2.5</sub> compositions.

To this end, in our study, we established a comprehensive framework for assessing air pollution exposure and its disparities in the urban context by leveraging big mobility data and street-level air quality measurements. The mobility dataset contains high spatiotemporal-resolution user locations from millions of unique mobile devices in the Bronx, New York City (NYC), an area that has a long history of being overburdened by public health issues related to environmental pollution<sup>29</sup>. The mobility layer was complemented by concurrent PM<sub>2.5</sub> data with street-segment-level granularity for estimating trajectory-based exposure as a function of location-specific pollutant concentration,



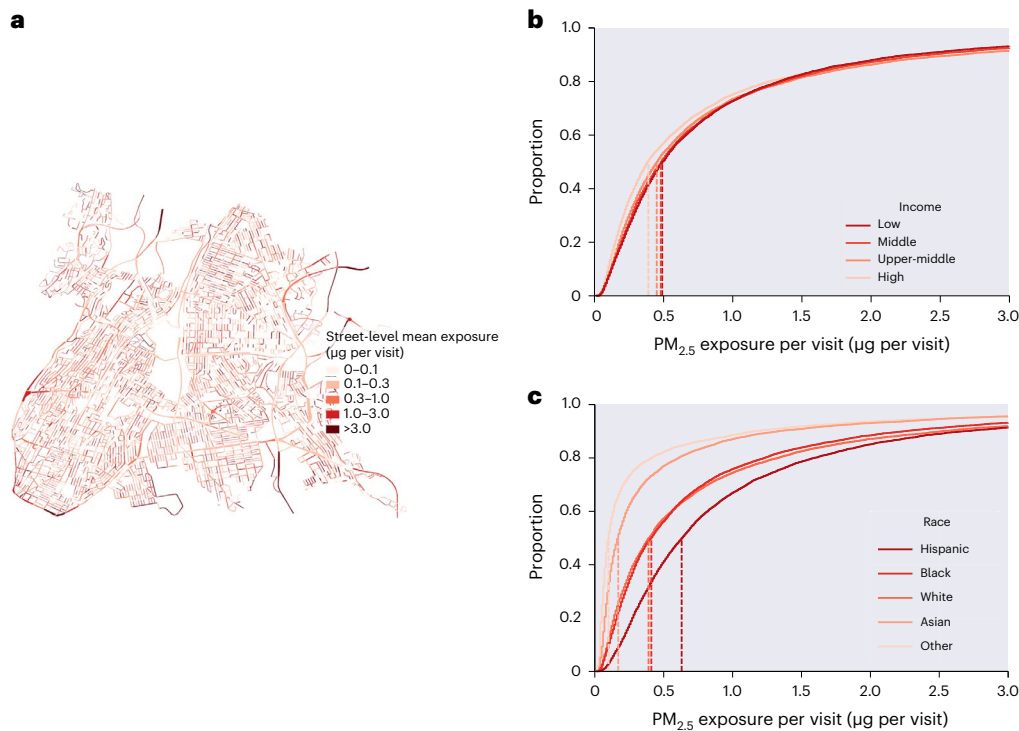
**Fig. 1 | Spatial distribution of predicted PM<sub>2.5</sub> concentrations and number of trips destined in the Bronx.** **a**, Predicted PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) in  $100 \times 100\text{-m}$  grid cells for average weather conditions in the Bronx. Daily predictions were made to pair with the mobility data in the analysis. **b**, Number of trips destined to the Bronx in  $200 \times 200\text{-m}$  grid cells without specifying income or race and/or ethnicity groups. Both maps reflect average conditions during the study period (September to November 2021).

dwelling time and inhalation rate. We characterized all Bronx visitors' street exposure, segment by segment, and clustered these data in different demographic groups, defined by the income level or the majority race and/or ethnicity at the participants' home locations, to quantify air pollution exposure disparities between and within the groups. A conceptual flowchart of our three-component data sources and methodology is presented in Supplementary Fig. 9. This study matches high-resolution mobility and air pollution data at a comparable level, and contextualizes hyperlocal air pollution exposure and its disparity in a dense and diverse urban area as a generalizable reference. This creates a population-representative 'panel' of individualized exposure tracking that is useful for evaluating public health, air pollution mitigation and urban design intervention outcomes.

## Results

### Air quality predictions and its descriptive link with people's mobility

People's mobility is largely different across income and race and/or ethnicity groups. We used people's mobility data from across NYC, but only accounted for their exposure in the Bronx, where there was detailed air pollution predictions and the highest racial and cultural diversity. We plotted a predicted surface of PM<sub>2.5</sub> concentrations in the study area, as shown in Fig. 1a, which was modeled using average weather conditions in autumn 2021. We only considered local PM<sub>2.5</sub>



**Fig. 2 | Street-level mean exposure in the Bronx and their cumulative distributions by income and race and/or ethnicity. a**, Average personal exposure ( $\mu\text{g}$  per visit) for all streets in the Bronx. **b, c**, Cumulative distribution function of  $\text{PM}_{2.5}$  exposure ( $\mu\text{g}$  per visit) when grouping people by income (**b**) and

race and/or ethnicity (**c**) by street. The function curves are colored from dark to light to denote high to low median personal exposure. The dashed lines indicate the median personal exposure for each group.

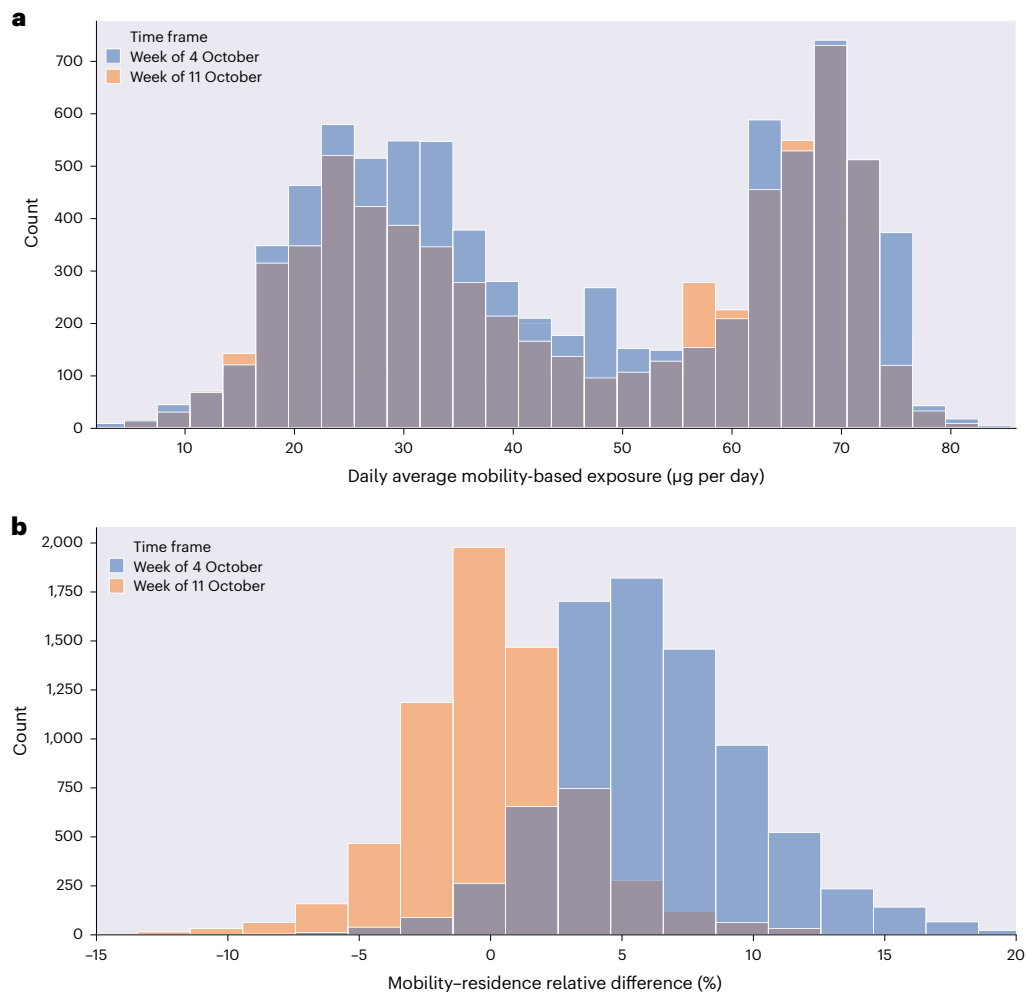
contributors with the background signals filtered. The average  $\text{PM}_{2.5}$  concentration, after background concentration filtering, was predicted to be  $4.36 \pm 0.33 \mu\text{g m}^{-3}$  (1 standard deviation). We observed several pollution hotspots in the South and East Bronx, as reported by the NYC government (<https://a816-dohbesp.nyc.gov/IndicatorPublic/data-features/nyccas/>). Our model highlights the importance of freight traffic in predicting  $\text{PM}_{2.5}$  pollution. The most notable pollution hotspot was in the southeastern Bronx (bottom right of Fig. 1a), where several highways intersect and high-pollution facilities are clustered. We used real-world observations and modeling with our results to exemplify how poor land use and infrastructure planning can lead to severe air pollution problems.

We were not able to directly display people's daily trajectories for data privacy reasons. Therefore, we employed trip origins and destinations inferred from individual trajectories as a surrogate for the contrast between people's mobility patterns and the air pollution distribution given in Fig. 1b. The trips were defined using the methodology in refs. 30,31. It should be noted that, in our exposure analysis, we still used individual trajectory data. All trips made from across NYC destined for the Bronx in the 3-month study period are plotted in the heat map (Fig. 1b). Compared to the air quality map, the mobility heat map has a coarser spatial resolution to satisfy the data provider's privacy requirements. More trips were made to the West Bronx, where the air pollution levels were lower than in the East or South Bronx, visitors to the latter areas thus being protected from high  $\text{PM}_{2.5}$  exposure. We also selected two consecutive weeks, from 4 to 17 October, in which to track daily individual exposure and its disparity. An evaluation of a longer period was not feasible because the data provider shuffled each individual's unique ID regularly to protect their privacy, and we were not informed exactly when this would be done. We picked individuals who both lived and moved around in the Bronx and who had at least 50 location records (7,000 individuals in total), and we calculated their daily average exposure. The first week (4–10 October)

presented a statistically significant difference between mobility and residence exposure range of  $6.29\% \pm 3.97\%$  when considering absolute values, with the residence-based method consistently estimating lower exposure levels. In the week of 11–17 October, there was a small relative difference between the residence and mobility exposures that was normally distributed, with a mean of  $0.09\% \pm 3.31\%$  (2.42%, on average, when considering absolute values of the relative difference). The breakdown by income and race and/or ethnicity groups produced extreme values among the Hispanic population, demonstrating the notable difference (of up to 20%) for this minority population. We illustrate the difference in exposure and its disparity in Supplementary Fig. 12.

### Air pollution exposure and its association with personal mobility

We perceive air prediction grid cells as merely arbitrary definitions for land-use regression, whereas street segments are links along which people move and which are increasingly considered to be public spaces rather than solely transportation infrastructure<sup>32,33</sup>. In this sense, street-level analyses can connect with planning practices to achieve 'desirable' public spaces<sup>34</sup> or to mitigate high emitters, such as truck traffic and commercial cooking<sup>35,36</sup>. Therefore, we first aggregated human exposure and its disparity to the street level to better review the intrinsic link between mobility and exposure among populations. We included all 7,500 street segments in the Bronx. On average, the street segments were  $163.78 \pm 125.63$  m (1 standard deviation) long and ranged from 27.50 to 2,460.00 m. Figure 2a shows that, on average, a person is exposed to  $1.20 \pm 3.22 \mu\text{g}$  (1 standard deviation) of  $\text{PM}_{2.5}$  on a Bronx street on a single visit. It should be noted that this distribution of personal exposure is highly skewed because some people work long hours or live on high-pollution streets. The maximal value of street-level personal exposure was  $45.25 \mu\text{g}$  per visit, in the southeastern Bronx, adjacent to the pollution hotspot we highlighted previously. The average exposure value is low because many people pass along



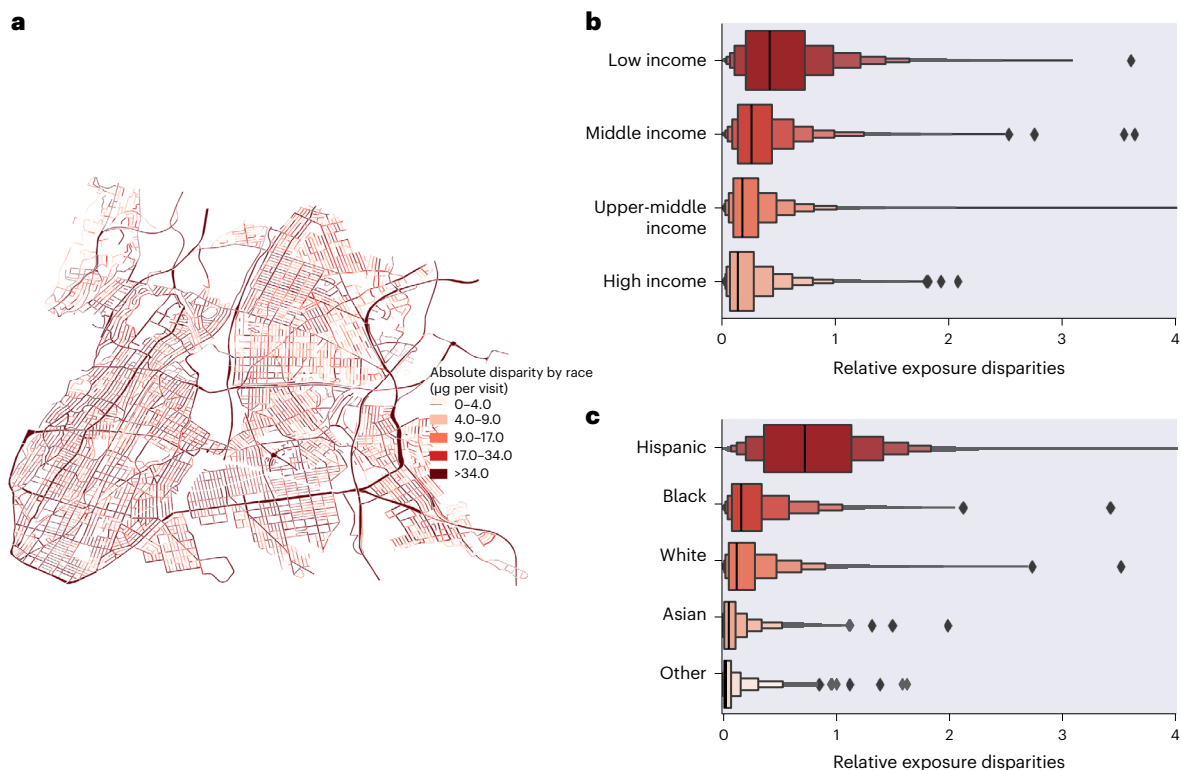
**Fig. 3 | The comparison between mobility- and residence-based exposure. a**, Daily average mobility-based exposure ( $\mu\text{g}$  per day) for the weeks of 4 October and 11 October. **b**, Relative difference between mobility- and residence-based exposure calculated as the difference between them divided by mobility-based exposure, where a positive value indicates the mobility-based exposure is higher.

streets without stopping. We conducted a health outcome analysis, in terms of attributable cardiovascular disease death, to contextualize our daily  $\text{PM}_{2.5}$  exposure levels (Supplementary Fig. 14 in Supplementary Information Section 5).

We further evaluated the street-level personal  $\text{PM}_{2.5}$  exposure by income and race and/or ethnicity groups. In Fig. 2a–c, we present the personal exposure for each Bronx street considering income and race and/or ethnicity. It should be noted that, in the following analysis, we report medians instead of averages to exclude extreme cases that might have occurred in any group, which could lead to skewed observations when compared with other groups. By ranking people's street-level exposure from low to high, in groups, it was revealed that the high-income group had been exposed to much less  $\text{PM}_{2.5}$  on the streets of the Bronx, with a median exposure of 0.38  $\mu\text{g}$  per visit compared to 0.45, 0.48 and 0.49  $\mu\text{g}$  per visit for the upper-middle-, middle- and low-income groups, respectively. With street-level exposure being negatively related to income, the absolute differences between the income groups were negligible. By contrast, by dividing people based on race- and/or ethnic-majority communities, differences in the street-level exposure between groups were evident. We noted that people from Hispanic-majority communities were much more exposed to  $\text{PM}_{2.5}$  (0.63  $\mu\text{g}$  per visit) than any other group. People from Black and white communities followed, with median exposures of 0.41 and 0.39  $\mu\text{g}$  per visit for each street segment. People from other race- and/or ethnic-majority

communities were exposed to the least  $\text{PM}_{2.5}$  (0.10  $\mu\text{g}$  per visit). That is to say, 50% of the people from Hispanic-majority communities experienced worse  $\text{PM}_{2.5}$  exposure than about 65% of the people from Black- or white-majority communities, 80% from Asian-majority communities and 85% from other-majority communities due to their mobility patterns. We employed absolute and relative disparity measures to further investigate the exposure disparity between and within these demographic groups.

In addition to the street-level analysis, we were also curious about how different the exposure was considering home location only and people's mobility patterns. We selected individuals who lived and move around in the Bronx, tracking their exposure for two typical weeks in October 2021. Figure 3 shows the daily average and relative differences between the residence and mobility exposure for the weeks of 4–10 October and 11–17 October. The nearly dichotomous distribution of mobility-based exposure signaled the mobility patterns of the residents who both worked and lived in the Bronx. The results showed that the residence- and mobility-based exposure were highly correlated for both weeks. The first week (4–10 October) presented a statistically significant difference between mobility and residence exposure, with a range of  $6.29\% \pm 3.97\%$  when considering absolute values, with the residence-based method consistently estimating lower exposure levels. In the second week (11–17 October), we observed a small relative difference between residence and mobility exposure that was normally



**Fig. 4 | Distributions of absolute and relative exposure disparities.** **a**, Street-level absolute disparity measure ( $\mu\text{g}$  per visit) for all Bronx streets by grouping people's majority neighbourhood race and/or ethnicity. **b, c**, The relative

disparity measure (Theil index) for all Bronx streets was calculated by grouping people's income (**b**) and race and/or ethnicity (**c**). The boxes in each demographic group show deciles of the within-group component of the Theil index.

distributed, with a mean of  $0.09\% \pm 3.31\%$  (2.42%, on average, when considering the absolute values of the relative difference). The breakdown by income and race and/or ethnic groups showed extreme values only among the Hispanic population, demonstrating a more notable difference (of up to 20%) for this minority population. Our findings are in line with those of previous studies<sup>16–18</sup> that estimated a 1% to 3% residence-to-mobility difference over the long term using travel surveys or cellular tower data. It is necessary to consider mobility data in such studies to be able to create refined exposure disparity mitigation policies that can address not only the total amount of  $\text{PM}_{2.5}$  exposure but also the location where the exposure happens. We observed that the residence-based absolute exposure disparity by majority race and/or ethnicity neighborhood was  $10.14 \mu\text{g}$  per day—higher than the mobility-based disparity ( $9.08 \mu\text{g}$  per day) over the two weeks—which indicates a risk of disparity overestimation when leaving people's mobility out. Exposure by income group showed virtually no difference, at  $4.97$  and  $4.99 \mu\text{g}$  per day, respectively. Obviously, exposure disparity by majority race and/or ethnicity was much higher than that by income group. However, our results are limited to the Bronx area and exposure in the 'out-of-home' time.

### Spatial exposure disparities between and within groups

We established the spatial disparity profiles of people physically present in the Bronx in terms of their income and race and/or ethnic community. The absolute disparity was calculated as the sum of the differences between a group's average exposure and the population-weighted average of all groups on the street. The same street visitors could be grouped either by four income levels or five races and/or ethnicities, with each street segment having one absolute disparity measuring the income dimension and the other measuring the race and ethnicity dimension. That is, we calculated the absolute disparity for the same population twice, but using two different grouping

criteria. Figure 4a illustrates the absolute disparity measures by race and/or ethnicity across the streets of the Bronx, the spatial patterns being very similar to those by income (Supplementary Fig. 13 in Supplementary Information Section 4). By plotting the logarithm of the exposure disparities by income and race and/or ethnicity for the same street against each other in Supplementary Fig. 13b, we showed that the exposure disparities were much more obvious when grouping the same population by race and/or ethnic community than by their income level, which confirms our earlier finding that race and/or ethnicity is a more important indicator of exposure disparity than income.

The highest exposure disparities, with respect to income and race and/or ethnicity, clustered around regional road arteries. We noted that the streets with high exposure disparities did not coincide with those with high average personal exposure, and only partially coincided with those with high air pollution levels, which suggests that existing methods for regulating air quality and lowering personal exposure might not mitigate exposure disparities between income or race and ethnic groups. We further evaluated the exposure disparities using the decomposable Theil index (relative exposure disparity), which is calculated as the sum of the within- and between-group disparities for each street. We observed that three-quarters of the disparities came from within each income or race and ethnic group, rather than from between groups. Figure 4b,c shows that, in addition to the absolute exposure, people from low-income and Hispanic-majority communities experienced the most within-group variation in  $\text{PM}_{2.5}$  exposure. The highest within-group disparity was found in people from Hispanic-majority communities, whereas, in terms of income, the highest disparity was found in the upper-middle income group.

### Discussion

From a hyperlocal perspective, the personal exposure levels and exposure disparities showed distinct spatial patterns. The New York City

Community Air Survey expanded its air monitoring network to high-poverty neighborhoods in 2014 by adding 15 environmental justice sites. The sites were chosen in high-poverty neighborhoods with a lower-than-average density of air quality monitors. We geo-matched the four Bronx sites with our street segments to allocate each site an absolute and a relative disparity measure (Supplementary Fig. 10 and Supplementary Table 3 in Supplementary Information Section 2). We observed two out of four sites located on streets with high exposure disparities in terms of income and race and/or ethnicity (the other two sites had low exposure disparities). This highlights the importance of identifying environmental justice hotspots from an exposure perspective that is more fundamentally related to population health. We further argue that equitable exposure mitigation plans should be reframed and go beyond air quality improvements in high-vulnerability neighborhoods to regulate emissions sources near points of interest that attract visitors from vulnerable communities.

A key takeaway from our study is that people from specific racial and ethnic-majority communities are disproportionately exposed to local air pollution. In addition, this phenomenon is more obvious than evaluating the same population by income. Similar results were observed by Jbaily et al.<sup>1</sup>, who found that, at zip-code-level granularity in the United States, there was a small difference (4%) between average PM<sub>2.5</sub> exposure levels in low- and high-income groups, but that the exposure levels between racial and ethnic groups were also apparent. This evidence suggests that not only is race and/or ethnicity a better exposure disparity indicator than income, but also the induced exposure disparity is systemic. Going one step further, our study proves quantitatively that this phenomenon is still present at a street-segment level, taking personal mobility into account. These findings suggest the need for targeted exposure mitigation policies to minimize air pollution exposure disparities in addition to generic air quality improvements. For example, it has been demonstrated that location-specific emissions reduction was much more effective than overall emissions reduction in reducing PM<sub>2.5</sub> exposure disparities among racial and/or ethnic groups in the United States<sup>21</sup>. This strategy's effectiveness would be even better at achieving a greater reduction in exposure disparities. Wang et al.<sup>37</sup> explored location-specific strategies to address photochemical air pollution in central California, pointing out that high-priority emissions sources differ at the city and regional scales. With a local scope, Kheirbek et al.<sup>38</sup> highlighted the importance of controlling truck and bus traffic to reduce exposure and its disparities among low-income neighborhoods. We explore exposure disparity by type of place in Supplementary Information Section 6, in which we overlapped the point-of-interest map and street-level exposure disparity, revealing that commercial places experience the highest disparities.

Although previous studies have estimated population-based exposure, they have not quantified within-group exposure disparity at the individual level, which has left a wide knowledge gap to be filled. Our personal mobility-based exposure analysis has demonstrated that the within-group exposure difference contributes up to three times more disparity than the between-group exposure difference. This phenomenon is more prominent for people from Hispanic-majority communities. This further demonstrates the double burden of NYC Hispanic-majority communities, in terms of their high absolute PM<sub>2.5</sub> exposure and the diverse exposure profiles of their individuals. This necessitates the consideration of personal mobility and hyperlocal air quality prediction in exposure disparity studies and practices, and a call for mitigation plans to address refined exposure and its systemic disparity for distinct demographic groups.

We recognize several limitations in our study. First, the study period was set in autumn 2021, during the COVID-19 pandemic. Although, during the observation period of the study, NYC had begun to come out of full lockdown and other restrictions on mobility, we acknowledge the pandemic's impacts on people's mobility patterns, which could have trickled downstream and affected our exposure and

disparity analysis. We cannot quantitatively determine the extent to which the same exposure patterns would hold true in the Bronx in a time with no limitations caused by pandemic restrictions. However, our exposure modeling reflects the exposure disparity caused by the combined factors of pandemic-induced mobility patterns and air quality changes, with previous literature having documented that the change in mobility patterns during the pandemic did not favor socially disadvantaged groups in terms of their exposure to COVID-19 infection and air pollution<sup>39–41</sup>. This is attributable to the fact that socially disadvantaged groups might have had no choice but to commute and attend work in person, and in heavily polluted areas<sup>40</sup>. Second, in our study, we did not differentiate PM<sub>2.5</sub> concentrations across travel modes, nor between indoor and outdoor environments. We also did not consider whether a person was wearing a mask or not during the pandemic. The assumptions we made have likely led to higher exposure estimates and smaller absolute exposure disparities in all communities, considering the better air quality during the pandemic<sup>42,43</sup> and the preventive mask-wearing<sup>44,45</sup>. Lastly, our air quality model provides daily predictions without including diurnal changes. The diurnal pattern of PM<sub>2.5</sub> is as evident as its spatial variation in urban areas, not only in terms of overall concentrations<sup>46,47</sup> but also their chemical compositions<sup>23,24</sup>, due to the operational patterns of various emissions sources. However, the temporal variability was barely able to be captured by our land-use regression-based air predictions, especially when we tried to maintain the model's robustness. Observations of PM<sub>2.5</sub> have to be averaged over unique days to reflect the most representative pollution state. A feasible solution would be to employ an atmospheric-chemistry-enabled physical model for time-step-by-time-step air quality prediction. Further studies are also encouraged to scale up our method and explore personal mobility-related exposure and its disparities in other urban areas. How post-pandemic mobility patterns might affect exposure disparities in the long run is another important topic to investigate.

## Methods

### Big mobility data sourcing and processing

The big mobility data was obtained through the Data for Good program by Cuebq. The dataset consisted of anonymized mobility data obtained from the smartphone location records of users who opted to share anonymous data through a California Consumer Privacy Act and General Data Protection Regulation-compliant framework. As part of Cuebq's Sensitive Points of Interest Policy, visits to privacy-sensitive locations were removed by the data provider. All researchers were contractually obliged not to share disaggregated data or results, or identify any individual users. Our use of the data complied with all relevant ethical regulations set by the data provider and by our universities. Our dataset consisted of location data records from ~500,000 unique phone users generating, on average, 500 million location records daily from NYC. The dataset covered a statistically significant fraction of NYC's residents<sup>31,48</sup> for the 3-month study period. The frequency of data point generation was dependent on how frequently a user was on their phone. We sampled a typical day to evaluate the average accuracy of the mobility dataset, which was  $19.61 \pm 18.29$  m (1 standard deviation), given the dense urban environment in the Bronx. Very rarely, there was an accuracy of more than 100 m, which will have had an actual effect on our exposure analysis. The location data sensitivity is discussed in Supplementary Information Section 1.3.

We selected so-called super-users, who generated at least 200 data points per day, to ensure robust exposure estimation and control for users who only stayed for a very short time in the study area, which comprised 20% of the total data. This approach had previously been adopted by Moro et al.<sup>31</sup>, who proved that users' mobility patterns were still representative of the population. We also tested the representativeness and sensitivity of our mobility data in Supplementary Information Section 1.1. To infer a home location and the presence of anonymous users from across NYC that were only visiting the Bronx, each user's

home location was determined considering three variables—number of days spent at a given location in the last month, daily average number of hours spent at that location and time of the day spent at the location (night time/daytime). A score was computed for each location of a unique device by combining these three variables. The more days and the higher the average number of hours the device spent at a location, especially at night, the higher the score. The location with the highest score was selected as the device's home location. One-time visitors to NYC were therefore filtered out. The algorithm ran daily to recursively update the inference, and only considered users with two or more days of data. Subsequently, the inferred home location was reassigned to the centroid of the Census Block Group (CBG) to protect the user's privacy. Given the large amount of data involved, and the iterative way the home location was inferred, the reported confidence of home inference was high (>99%).

Based on the inferred home location, each unique user was assigned to a CBG matching demographic data from the American Decennial Census 2020 and the 5-year American Community Survey 2020 (<http://www.census.gov/programs-surveys/acs/data.html>). People's economic status was determined from their annual income level in quartiles from low (below US\$67,000), to medium-low (between US\$67,000 and US\$89,999), medium-high (between US\$90,000 and US\$113,999) and high (equal to or above US\$114,000) income<sup>31</sup>. Regarding the definitions of race and ethnicity, we first divided the majority group in each CBG into five categories—Latino and/or Hispanic as Hispanic, non-Hispanic as different races, including Asian, Black, white and other, for Native American, Pacific Islander, multiracial or undefined<sup>4,8,21</sup>. The definition of each CBG's majority race and/or ethnicity was purely based on the population, with no lower bound of the threshold in the population threshold. For each individual's trajectory, we did not infer their race and/or ethnicity because this might have led to the stereotyping of the mobility patterns of a certain race and/or ethnicity. We also did not want to lose the difference in latent mobility pattern. Therefore, each user's trajectory was tagged as 'from an Asian-majority CBG', 'from a Black-majority CBG', 'from a Hispanic-majority CBG', 'from a white-majority CBG' or 'from an other-majority CBG', determined by the race and/or ethnicity with the highest population percentage in their home CBG. We were aware of the potential risk of miscategorizing individual race and ethnicity trajectories, and paid special attention to interpreting the related results, given the extreme cases where two or more racial and/or ethnic groups had similar populations in a CBG. However, we also did not want to arbitrarily set a population difference threshold because this could have led to the underrepresentation of certain communities. This manipulation was the best we could apply to minimize stereotypical presumptions based on race- and/or ethnicity-related mobility patterns without it affecting our conclusions.

### Hyperlocal air quality data collection and modeling

We collected hyperlocal air quality data—in this context, PM<sub>2.5</sub> mass concentrations—using a fleet of Internet of Things-enabled environmental sensing platforms<sup>49</sup>. The sensing platform was equipped with an optical particle counter, a temperature and humidity sensor, a high-accuracy global positioning system, and a control board for real-time sensor synchronization, coordination and data upload to the cloud. Self-powered by a solar panel and battery pack, the platform was able to run independently once deployed on top of a vehicle. Five sensing platforms were mounted on civic services vehicles operated by the Department of Parks and Recreation, NYC. In the 3-month data collection campaign, these platforms circulated around the Bronx on all days during the daytime, collecting data every 5 s, collecting over 600,000 valid data points. Given that we were using non-reference monitors, we developed a rigorous framework for sensor calibration and data quality assurance<sup>50</sup>, as documented in Supplementary Information Section 2.

The 3-month air quality monitoring campaign expanded over one-third of the Bronx's major roads. We trained an empirical model for air

quality prediction to extrapolate the air quality observations to a full land coverage of the Bronx, following the general land-use regression procedures adopted by Messier et al.<sup>51</sup>. First, the PM<sub>2.5</sub> records were corrected to their regional transport components to focus on local emissions sources and points of interest. Then, the corrected geo-located PM<sub>2.5</sub> observations were aggregated to 100 × 100-m grid cells, forming 1,213 cells with valid data records. Each grid cell contained at least 10 unique observations, collected on a minimum of four distinct days, to represent a typical air quality profile of the cell. The mean PM<sub>2.5</sub> concentration in each grid cell was treated as the target variable, and a series of land-use and emissions source variables were used as explanatory variables, following the same method used by the New York City Community Air Survey (<https://a816-dohbhesp.nyc.gov/IndicatorPublic/data-features/nyccas/>). We adopted a gradient-boosting tree model—the Light Gradient Boosting Machine (LightGBM)<sup>52</sup>—for air pollution level prediction. Lastly, the air pollution level on each street segment was estimated as the weighted average concentration by length, if a street segment spanned across several cells. It should be noted that air quality prediction maps were generated daily from 1 September to 30 November, in synchronization with the mobility data, considering only the changes in meteorology (for example, temperature, relative humidity, wind speed and air pressure). We describe the data calibration, aggregation, regression modeling and validation in detail in Supplementary Information Section 3. The air prediction model's performance, and its comparison with the reference monitoring, are documented in Supplementary Information Section 1.2.

### Exposure definition and calculation

**Street-level exposure calculation.** Individual exposure estimation provides key information for the risk assessment and health analysis of evidence-based emissions regulations and public health policymaking. Exposure to a contaminant—in our case PM<sub>2.5</sub>—is defined as an event that occurs when there is contact at a boundary between a human and the environment with a specific PM<sub>2.5</sub> concentration for an interval of time<sup>53</sup>. Previous studies analyzing air pollution disparities have heavily depended on population-weighted concentrations, rather than individual exposure assessments that take individual mobility patterns into account. Given that our hyperlocal air quality and human mobility data were well coupled in temporal and spatial granularity, we first developed our exposure calculation methods at the street-segment level. The street segments were generated by creating flat-end buffers of different sizes, referencing the Open Street Map 'ways' elements, often referred to as street segments. We fragmented the street segments longer than two blocks to capture the variability in PM<sub>2.5</sub> concentrations. The street segments were confined by connections to other street segments (intersections). The boundaries of the street segments were drawn based on the most recent satellite images.

Based on the US Environmental Protection Agency's air pollutants exposure model (<https://www.epa.gov/fera/human-exposure-modeling-air-pollutants-exposure-model>), our exposure calculation per visit to a street, which can be aggregated by demographic groups and over time, is presented in the following equation. The street-segment aggregation provided a subtle normalization of the various amounts of time people spend in the Bronx. In addition, the aggregation was flexible in considering the air quality and mobility data at different spatial scales. For instance, in our case, we considered mobility data from all over NYC and the exposure in the Bronx.

$$e_j = \sum_{i \in j} t_{i,j} \times C_{PM} \times I_{HR} \quad (1)$$

Where  $e_j$  is the exposure in  $\mu\text{g}$  per visit for demographic group  $j$  on a street segment;  $t_{i,j}$  is the time that individual  $i$  that belongs to demographic group  $j$  spent on the street segment;  $C_{PM}$  is the PM<sub>2.5</sub> concentration on the street segment in  $\mu\text{g m}^{-3}$ ; and  $I_{HR}$  is the inhalation rate of an

average adult, which is constant at  $0.012 \text{ m}^{-3} \text{ min}^{-1}$  (<https://www.epa.gov/expobox/exposure-factors-handbook-chapter-6>).

It should be noted that, technically, equation (1) calculates the mass of inhaled  $\text{PM}_{2.5}$ . However, because the concepts of ‘exposure’ and ‘dose’ are often misused, we followed the definitions from ref. 53 and instead opted for the term ‘exposure’ to refer to the outcome of equation (1) for better communication with a broader audience. Although the inhalation rate in equation (1) is merely a constant, it generates more understandable results in the unit of inhaled  $\text{PM}_{2.5}$  mass (that is,  $\mu\text{g}$  per visit), allowing better characterization of population exposure disparities in the Bronx. Our exposure calculation bridges population exposure estimation and individual exposure tracking. In the presence of high-resolution mobility data, we were able to model spatiotemporally resolved individual exposure that was representative of the population. For  $\text{PM}_{2.5}$ , the terms exposure and dose are often used interchangeably, given its high penetration rate across the upper respiratory tract.

**Residence- and mobility-based exposure calculation.** To contrast the residence- and mobility-based exposure, we only chose individuals who both lived and moved around in the Bronx, where we had detailed air quality predictions, who had at least 50 location records (7,000 individuals in total). We selected two typical weeks in October (4–10 October and 11–17 October) to demonstrate possible differences in their exposure calculations. Because the data provider shuffled individuals’ unique IDs from time to time (every few weeks to months) to preserve the data contributors’ privacy, we could not track the same individuals for the whole study period. This is a practical obstacle to analyzing exposure and its disparity at the individual level. For each individual, mobility-based exposure was computed following the same definition as in the original study—as the product of the street-specific  $\text{PM}_{2.5}$  values, the amount of time the individual spent on all the visited streets and the standard inhalation rate. To compare residence- and mobility-based exposure on the same quantitative scale, the residence-based air pollution exposure was attributed to each user by assigning the  $\text{PM}_{2.5}$  value corresponding to the CBG home location, multiplying it by the same on-the-street (out-of-home) time and the inhalation rate. It is worth mentioning that we adopted a distinct way of calculating the disparities, as all individuals/trajectories from the same CBG would have had the same sociodemographic information, and the disparities could only be calculated at the CBG level or coarser. The general idea was to treat the whole Bronx region as one big street segment, where we accounted for the absolute exposure disparities by dividing the 7,000 individuals into sociodemographic groups. We conducted a health outcome analysis by attributing cardiovascular disease mortality to ambient  $\text{PM}_{2.5}$  exposure (in  $\mu\text{g}$  per day), finding that there was a 16.66% increase in cardiovascular disease mortality attributable to  $\text{PM}_{2.5}$  in the 9th exposure decile (71.20  $\mu\text{g}$  per day) compared with the 1st decile (21.42  $\mu\text{g}$  per day). Details about this analysis have been added to Supplementary Information Section 5.

### Exposure disparity characterization

A myriad of disparity measures exist, quantifying the problem in many dimensions. The most common ones include the coefficient of variation<sup>1</sup> for air pollution exposure disparities, the Gini coefficient, information entropy, the Atkinson index and the income segregation measure for places<sup>31</sup>. We selected two disparity metrics that fit with this study’s context to measure street-segment-level air pollution exposure disparity.

First, we considered an absolute disparity measure (hereafter, the absolute measure), which provided a simple physical explanation of exposure disparity. It summed up the absolute  $\text{PM}_{2.5}$  exposure difference between each group and the population-weighted average exposure level on each street segment. The absolute measure ranges from 0 to positive infinity, where 0 denotes perfect equality, in which case all

groups share the same level of  $\text{PM}_{2.5}$  exposure. The more deviated each group is from the average exposure, the higher the absolute measure is on the street. The following equation details the absolute measure:

$$El_{\text{abs}} = \sum_j |\bar{e}_j - \mu| \quad (2)$$

Where  $El_{\text{abs}}$  is the absolute exposure disparity measure in  $\mu\text{g}$  per visit as the summation of absolute differences on each street segment;  $\bar{e}_j$  is the average exposure of group  $j$ ; and  $\mu$  is the population-weighted average exposure of people in all demographic groups.

We also wanted to quantify the disparity using a normalized metric, where we could compare street to street on an equal footing. Theil’s  $T$  index was employed to investigate exposure disparity on each street, which is essentially a special case of the generalized entropy index, measuring the disparities of air pollution exposure in a population. As demonstrated in equation (4), the Theil index of all groups on the street is the summation of within-group disparity (the first term) and between-group disparity (the second term). The within-group disparity is calculated as the weighted-averaged Theil indices of each demographic group, where each person in a group is an agent for calculation, as in equation (3). The between-group disparity equates to the Theil index treating each group as an agent. While the Theil index is always positive, its decomposability allows us to evaluate each group’s contribution to the disparity. The index can be positive, indicating more exposure was experienced by a group proportionate to its population, or it can be negative, with less exposure experienced by a group.

$$T_j = \frac{1}{N_j} \sum_{i \in j} \frac{e_i}{\bar{e}_j} \ln \frac{e_i}{\bar{e}_j} \quad (3)$$

$$T_T = \sum_j s_j T_j + \sum_j s_j \ln \frac{\bar{e}_j}{\mu} \text{ for } s_j = \frac{N_j \bar{e}_j}{N \mu} \quad (4)$$

Where  $T_j$  is the Theil index for a demographic group  $j$  on a street segment;  $N_j$  is the total number of people in group  $j$ ;  $e_i$  is person  $i$ ’s exposure, who belongs to group  $j$ ;  $\bar{e}_j$  is the average exposure of group  $j$ ;  $T_T$  is the Theil index considering all demographic groups;  $s_j$  is the proportion of exposure experienced by group  $j$ ;  $\mu$  is the average exposure of people in all demographic groups; and  $N$  is the total number of people from all demographic groups.

### Reporting summary

Further information on the research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The air quality predictions, exposure by street segment, exposure by individual trajectory and exposure disparity measures data are publicly available via Zenodo at <https://doi.org/10.5281/zenodo.11044847> (ref. 54). The personal mobility data were obtained from Cuebiq through their social impact and Data for Good program. They were analyzed under a strict data privacy agreement between the authors and the company and, therefore, are not publicly available. Interested parties can request access to this mobility dataset, with more information, from <https://spectus.ai/social-impact/>.

### References

1. Jbaily, A. et al. Air pollution exposure disparities across US population and income groups. *Nature* **601**, 228–233 (2022).
2. Kim, S. Y. et al. Concentrations of criteria pollutants in the contiguous U.S., 1979 – 2015: role of prediction model parsimony in integrated empirical geographic regression. *PLoS ONE* **15**, e0228535 (2020).

3. Fann, N., Kim, S. Y., Olives, C. & Sheppard, L. Estimated changes in life expectancy and adult mortality resulting from declining PM<sub>2.5</sub> exposures in the contiguous United States: 1980–2010. *Environ. Health Perspect.* **125**, 097003 (2017).
4. Chambliss, S. E. et al. Local- and regional-scale racial and ethnic disparities in air pollution determined by long-term mobile monitoring. *Proc. Natl Acad. Sci. USA* **118**, e2109249118 (2021).
5. Liu, J. et al. Disparities in air pollution exposure in the United States by race/ethnicity and income, 1990–2010. *Environ. Health Perspect.* **129**, 127005 (2021).
6. Clark, L. P., Millet, D. B. & Marshall, J. D. National patterns in environmental injustice and inequality: outdoor NO<sub>2</sub> air pollution in the United States. *PLoS ONE* **9**, e94431 (2014).
7. Clark, L. P., Millet, D. B. & Marshall, J. D. Changes in transportation-related air pollution exposures by race-ethnicity and socio-economic status: outdoor nitrogen dioxide in the United States in 2000 and 2010. *Environ. Health Perspect.* **125**, 097012 (2017).
8. Tessum, C. W. et al. PM<sub>2.5</sub> pollutants disproportionately and systemically affect people of color in the United States. *Sci. Adv.* **7**, 4491–4519 (2021).
9. Brazil, N. Environmental inequality in the neighborhood networks of urban mobility in US cities. *Proc. Natl Acad. Sci. USA* **119**, e2117776119 (2022).
10. Lane, H. M., Morello-Frosch, R., Marshall, J. D. & Apte, J. S. Historical redlining is associated with present-day air pollution disparities in U.S. cities. *Environ. Sci. Technol. Lett.* **9**, 345–350 (2022).
11. Schell, C. J. et al. The ecological and evolutionary consequences of systemic racism in urban environments. *Science (1979)* **369**, eaay4497 (2020).
12. Crouse, D. L., Ross, N. A. & Goldberg, M. S. Double burden of deprivation and high concentrations of ambient air pollution at the neighbourhood scale in Montreal, Canada. *Soc. Sci. Med.* **69**, 971–981 (2009).
13. Alessandretti, L., Aslak, U. & Lehmann, S. The scales of human mobility. *Nature* **587**, 402–407 (2020).
14. Pappalardo, L. et al. Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**, 8166 (2015).
15. Kwan, M. P. The neighborhood effect averaging problem (NEAP): an elusive confounder of the neighborhood effect. *Int. J. Environ. Res. Public Health* **15**, 1841 (2018).
16. de Souza, P. et al. Quantifying disparities in air pollution exposures across the United States using home and work addresses. *Environ. Sci. Technol.* **58**, 280–290 (2023).
17. Reis, S. et al. The influence of residential and workday population mobility on exposure to air pollution in the UK. *Environ. Int.* **121**, 803–813 (2018).
18. Nyhan, M. M. et al. Quantifying population exposure to air pollution using individual mobility patterns inferred from mobile phone data. *J. Expo. Sci. Environ. Epidemiol.* **29**, 238–247 (2019).
19. Nyhan, M. et al. ‘Exposure track’ - The impact of mobile-device-based mobility patterns on quantifying population exposure to air pollution. *Environ. Sci. Technol.* **50**, 9671–9681 (2016).
20. Shekarrizfard, M., Faghih-Imani, A. & Hatzopoulou, M. An examination of population exposure to traffic related air pollution: comparing spatially and temporally resolved estimates against long-term average exposures at the home location. *Environ. Res.* **147**, 435–444 (2016).
21. Wang, Y. et al. Location-specific strategies for eliminating US national racial-ethnic PM<sub>2.5</sub> exposure inequality. *Proc. Natl Acad. Sci. USA* **119**, e2205548119 (2022).
22. Manousakas, M. et al. Quantitative assessment of the variability in chemical profiles from source apportionment analysis of PM<sub>10</sub> and PM<sub>2.5</sub> at different sites within a large metropolitan area. *Environ. Res.* **192**, 110257 (2021).
23. Zhao, S. et al. Temporal variation characteristics and source apportionment of metal elements in PM<sub>2.5</sub> in urban Beijing during 2018–2019. *Environ. Pollut.* **268**, 115856 (2021).
24. Jeong, C. H. et al. Temporal and spatial variability of traffic-related PM<sub>2.5</sub> sources: comparison of exhaust and non-exhaust emissions. *Atmos. Environ.* **198**, 55–69 (2019).
25. Li, R. et al. Key toxic components and sources affecting oxidative potential of atmospheric particulate matter using interpretable machine learning: insights from fog episodes. *J. Hazard. Mater.* **465**, 133175 (2024).
26. Wu, D. et al. Achieving health-oriented air pollution control requires integrating unequal toxicities of industrial particles. *Nat. Commun.* **14**, 6491 (2023).
27. Cheng, K. et al. Life-course health risk assessment of PM<sub>2.5</sub> elements in China: exposure disparities by species, source, age, gender, and location. *Environ. Sci. Technol.* **58**, 3629–3640 (2024).
28. Jin, L. et al. Contributions of city-specific fine particulate matter (PM<sub>2.5</sub>) to differential in vitro oxidative stress and toxicity implications between Beijing and Guangzhou of China. *Environ. Sci. Technol.* **53**, 2881–2891 (2019).
29. Spira-Cohen, A., Chen, L. C., Kendall, M., Lall, R. & Thurston, G. D. Personal exposures to traffic-related air pollution and acute respiratory health among Bronx schoolchildren with asthma. *Environ. Health Perspect.* **119**, 559–565 (2011).
30. Hariharan, R. & Toyama, K. Project Lachesis: parsing and modeling location histories. Conference paper. *Lecture Notes in Computer Science* **3234**, 106–124 (2004).
31. Moro, E., Calacci, D., Dong, X. & Pentland, A. Mobility patterns are associated with experienced income segregation in large US cities. *Nat. Commun.* **12**, 4633 (2021).
32. von Schönfeld, K. C. & Bertolini, L. Urban streets: epitomes of planning challenges and opportunities at the interface of public space and mobility. *Cities* **68**, 48–55 (2017).
33. Karndacharuk, A., Wilson, D. J. & Dunn, R. A review of the evolution of shared (street) space concepts in urban environments. *Transp. Rev.* **34**, 190–220 (2014).
34. Salazar Miranda, A., Fan, Z., Duarte, F. & Ratti, C. Desirable streets: using deviations in pedestrian trajectories to measure the value of the built environment. *Comput. Environ. Urban Syst.* **86**, 101563 (2021).
35. Shukla, K. et al. ZIP code-level estimation of air quality and health risk due to particulate matter pollution in New York City. *Environ. Sci. Technol.* **56**, 7119–7130 (2022).
36. Johnson, S., Haney, J., Cairone, L., Huskey, C. & Kheirbek, I. Assessing air quality and public health benefits of New York City’s climate action plans. *Environ. Sci. Technol.* **54**, 9804–9813 (2020).
37. Wang, Y., Bastien, L., Jin, L. & Harley, A. R. Location-specific control of precursor emissions to mitigate photochemical air pollution. *Environ. Sci. Technol.* **57**, 9693–9701 (2023).
38. Kheirbek, I., Haney, J., Douglas, S., Ito, K. & Matte, T. The contribution of motor vehicle emissions to ambient fine particulate matter public health impacts in New York City: a health burden assessment. *Environmental Health* **15**, 1–14 (2016).
39. Kerr, G. H., Goldberg, D. L. & Anenberg, S. C. COVID-19 pandemic reveals persistent disparities in nitrogen dioxide pollution. *Proc. Natl Acad. Sci. USA* **118**, e202240918 (2021).
40. Huang, Y. & Li, R. The lockdown, mobility, and spatial health disparities in COVID-19 pandemic: a case study of New York City. *Cities* **122**, 103549 (2022).
41. Hong, B., Bonczak, B. J., Gupta, A., Thorpe, L. E. & Kontokosta, C. E. Exposure density and neighborhood disparities in COVID-19 infection risk. *Proc. Natl Acad. Sci. USA* **118**, e2021258118 (2021).

42. Hudda, N., Simon, M. C., Patton, A. P. & Durant, J. L. Reductions in traffic-related black carbon and ultrafine particle number concentrations in an urban neighborhood during the COVID-19 pandemic. *Sci. Total Environ.* **742**, 140931 (2020).
43. Perera, F. et al. Potential health benefits of sustained air quality improvements in New York City: a simulation based on air pollution levels during the COVID-19 shutdown. *Environ. Res.* **193**, 110555 (2021).
44. Laumbach, R. J. & Cromar, K. R. Personal interventions to reduce exposure to outdoor air pollution. *Annu. Rev. Public Health* **43**, 293–309 (2022).
45. Laumbach, R. J. et al. Personal interventions for reducing exposure and risk for outdoor air pollution. *Ann. Am. Thorac. Soc.* **18**, 1435–1443 (2021).
46. Lu, Y., Giuliano, G. & Habre, R. Estimating hourly PM<sub>2.5</sub> concentrations at the neighborhood scale using a low-cost air sensor network: a Los Angeles case study. *Environ. Res.* **195**, 110653 (2021).
47. Sun, J., Gong, J. & Zhou, J. Estimating hourly PM<sub>2.5</sub> concentrations in Beijing with satellite aerosol optical depth and a random forest approach. *Sci. Total Environ.* **762**, 144502 (2021).
48. Deng, H., Du, J., Gao, J. & Wang, Q. Network percolation reveals adaptive bridges of the mobility network response to COVID-19. *PLoS ONE* **16**, e0258868 (2021).
49. Mora, S., Anjomshoaa, A., Benson, T., Duarte, F. & Ratti, C. Towards large-scale drive-by sensing with multi-purpose city scanner nodes. In *2019 IEEE 5th world forum on Internet of Things (IEEE, 2019)*; <https://doi.org/10.1109/WF-IOT.2019.8767186>
50. Wang, A. et al. Leveraging machine learning algorithms to advance low-cost air sensor calibration in stationary and mobile settings. *Atmos. Environ.* **301**, 119692 (2023).
51. Messier, K. P. et al. Mapping air pollution with Google Street View cars: efficient approaches with mobile monitoring and land use regression. *Environ. Sci. Technol.* **52**, 12563–12572 (2018).
52. Ke, G. et al. LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **30**, 52 (2017).
53. Sexton, K. & Ryan, P. B. in *Air Pollution, the Automobile, and Public Health* (eds Watson, A. Y., Bates, R. R. & Kennedy, D.) 207–238 (National Academy Press, 1988).
54. Testi, I. Air Quality and Exposure Disparity Results for the Bronx, New York City. *Zenodo* <https://doi.org/10.5281/zenodo.11044847> (2024).

## Acknowledgements

This study was funded by the MIT Senseable City Lab Consortium (Dubai Future Foundation, Toyota, UnipolTech, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Volkswagen Group America, FAE Technology, MipMap, GoAigua, Shell, ENEL Foundation, Kyoto University, SMART – Singapore-MIT Alliance for Research and Technology, Weizmann Institute of Science, Universidad Autónoma de Occidente, Instituto Politecnico Nacional, Imperial College London,

Università di Pisa, KTH Royal Institute of Technology and the AMS Institute). The study was conducted with the support of the Center for Climate-Smart Transportation, Johns Hopkins University, the NYC Office of Technology and Innovation, NYC Department of Health, NYC Department of Citywide Administrative Services, and the Spectus Social Impact team. Specifically, we thank S. Johnson from the New York City Department of Health, and E. Ilten and B. Lake from Cuebq for their continuous help in air quality and mobility data acquisition and usage.

## Author contributions

I.T. was involved in the conceptualization, writing of the paper, methodology and data acquisition. A.W. was involved in conceptualization, writing of the paper, methodology, data acquisition and supervision. S.P. was involved in the writing of the paper. S.M. was involved in supervision and project management. E.W. was involved in the methodology and writing of the paper. M.N. was involved in the methodology and writing of the paper. F.D. was involved in writing of the paper and supervision. P.S. was involved in the methodology, writing of the paper and supervision. C.R. was involved in supervision and funding acquisition.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44284-024-00093-x>.

**Correspondence and requests for materials** should be addressed to An Wang or Simone Mora.

**Peer review information** *Nature Cities* thanks Yuzhou Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                      | Confirmed  |
|--------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Air quality predictions, exposure by street segment, exposure by individual trajectory, and exposure disparity measures data are publicly available in the data product folder at <https://github.com/MIT-Senseable-City-Lab/OSCS>. The personal mobility data are obtained from Spectus Inc. through their social impact and 'data

for good' program, which were analyzed under a strict data privacy agreement between the authors and the company and, therefore, are not publicly available. Interested parties can request access to this mobility dataset with more information from <https://spectus.ai/social-impact/>.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	We do not report on sex- or gender-based analysis as we do not have relevant data.
Reporting on race, ethnicity, or other socially relevant groupings	We group people based on race/ethnicity and income level based on the American Census data and American Community Survey given inferred individual home locations.
Population characteristics	See above
Recruitment	No recruitment is involved in this study.
Ethics oversight	Massachusetts Institute of Technology; Spectus Inc.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Our study establishes a comprehensive framework for assessing air pollution exposure and its disparities in the urban context by leveraging big mobility data and street-level air quality measurements.
Research sample	The mobility dataset contains high spatio-temporal resolution user locations from millions of unique mobile devices in the Bronx, New York City (NYC).
Sampling strategy	Our dataset consisted of location data records from ~500,000 unique phone users generating, on average, 500 million location records daily from New York City. The dataset covers a statistically significant fraction of New York City residents.
Data collection	The big mobility data was obtained through the 'Data for Good' program by Spectus Inc. The dataset consists of anonymized mobility data obtained from smartphone location records of users who opted-in to share anonymous data through a California Consumer Privacy Act and General Data Protection Regulation-compliant framework. All researchers were contractually obligated to not share disaggregated data and results or identify any individual users.
Timing	September to December, 2021
Data exclusions	No data were included from the study.
Non-participation	No participant dropped out the study.
Randomization	People are allocated into race/ethnicity and income groups based on their home location census block group.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Plants

## Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

## Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

## Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.