nature sustainability

ARTICLES

https://doi.org/10.1038/s41893-022-00856-1

Disparate air pollution reductions during California's COVID-19 economic shutdown

Richard Bluhm[1,2,8], Pascal Polonik[3,8], Kyle S. Hemes[4,8], Luke C. Sanford[2,5,6,8], Susanne A. Benz[6,7,8], Morgan C. Levy[3,6,8], Katharine L. Ricke[3,6] and Jennifer A. Burney[6] ✉

Minority communities in the United States often experience higher-than-average exposures to air pollution. However, the relative contribution of institutional biases to these disparities can be difficult to disentangle from other factors. Here, we use the economic shutdown associated with the 2020 COVID-19 shelter-in-place orders to causally estimate pollution exposure disparities caused by the in-person economy in California. Using public and citizen-science ground-based monitor networks for respirable particulate matter, along with satellite records of nitrogen dioxide, we show that sheltering in place produced disproportionate air pollution reductions for non-White (especially Hispanic and Asian) and low-income communities. We demonstrate that these racial and ethnic effects cannot be explained by weather patterns, geography, income or local economic activity as measured by local changes in mobility. They are instead driven by regional economic activity, which produces local harms for diffuse economic benefits. This study thus provides indirect, yet substantial, evidence of systemic racial and ethnic bias in the generation and control of pollution from the portion of the economy most impacted in the early pandemic period.

There exist substantial concerns in the United States about the pervasive harms of racism, which modern scholarship conceptualizes as either active or passive normalization of racial or ethnic inequities[1]. (In this Article, we use 'racism' in the modern descriptive sense that does not hinge on the intent of the perpetrator(s); that is, actions and policies that promote race-based inequities are racist, whether or not such an outcome is intended.) Particularly worrisome is the potential for institutionalized (or systemic) racism—in the form of policies, regulations and norms that favour certain racial or ethnic groups[2]—to perpetuate such harm via democratic processes. Rigorous quantitative evidence of institutional racism can be difficult to come by because the effects of various social and institutional processes that may embed bias (for example, urban planning and environmental regulation) often overlap in space and time and thus stymie attempts at more specific attribution (for example, ref. [3]). This in turn makes policy proposals that address racism head-on more difficult to justify. This has long been the case with environmental injustice, or the manifestation of systemic racism in environmental policymaking and enforcement[4]. (The term 'environmental injustice' is often used more broadly to describe disparities across multiple demographic dimensions, including, but not limited to, race and ethnicity. In this Article, for clarity, we use this more specific definition with a primary focus on racial and ethnic bias in environmental policy.)

Disparities in air pollution concentrations provide a clear example of this attribution problem[5,6]. Air pollution is linked to a wide range of negative health consequences[7] and is estimated to cause nearly 9 million premature deaths globally per year[8]. On average, these health effects are not distributed evenly among different demographic groups[9–11], running counter to the notion that society's environmental burdens should be equally shared[5,6,12]. However, despite observable exposure gradients across racial and ethnic groups, causally ascribing such inequities to bias in environmental policy has proved difficult. (In this Article, we consider 'environmental policy' to be the full landscape of policies, laws, statutes, regulations and enforcement mechanisms governing environmental quality. This definition includes gaps; that is, existing loopholes, lack of regulation and non-enforcement of rules are also forms of policy.) Economic and other social policymaking (for example, housing, transportation, education) over generations has created the modern geography of who lives where. Over time, myriad physical and social confounds—including variable atmospheric transport processes[13], economic inequalities[10,14] and neighbourhood demographics[9,15]—have become correlated with present-day pollution exposures. As such, moving beyond simple observations of disparate but confounded exposures in contemporary cross sections to causal attribution of environmental injustice requires additional evidence. This can be achieved by a random perturbation to the status quo[16]; such a shock to the policy regime was provided by the initial COVID-19 economic shutdown in California[17].

In early 2020, governments implemented unprecedented policies to limit the public health impacts of the COVID-19 pandemic, including stay-at-home orders and travel limitations, with California instituting some of the most aggressive lock-down measures in the United States[18]. The well-known side effect of these policies was widespread economic shutdown: businesses closed, factories shuttered and employees temporarily discontinued their daily commutes[19]. (In California, 60.51% of businesses reported a decline in demand, 22.27% reported closure due to government mandate and only 13.9% indicated that the pandemic had no impact on their business[20]; the transport sector accounted for an estimated 97.5% of the decline in CO_2 emissions in spring 2020 over the same period in 2019[21].) Because pollutants such as particulate matter with diameter smaller than 2.5 μm ($PM_{2.5}$) and nitrogen

[1]School of Economics and Management, Leibniz University Hannover, Hanover, Germany. [2]Department of Political Science, UC San Diego, San Diego, CA, USA. [3]Scripps Institution of Oceanography, UC San Diego, San Diego, CA, USA. [4]Woods Institute for the Environment, Stanford University, Stanford, CA, USA. [5]School of the Environment, Yale University, New Haven, CT, USA. [6]School of Global Policy and Strategy, UC San Diego, San Diego, CA, USA. [7]Centre for Water Resources Studies, Dalhousie University, Halifax, Nova Scotia, Canada. [8]These authors contributed equally: Richard Bluhm, Pascal Polonik, Kyle S. Hemes, Luke C. Sanford, Susanne A. Benz, Morgan C. Levy. ✉e-mail: jburney@ucsd.edu

dioxide (NO_2) are produced by transportation, industrial processes, energy production and agriculture[22], pollutant concentrations tend to track aggregate economic activity[23,24]. The lock-down corresponded to reductions in both satellite and ground-based observations of NO_2 and $PM_{2.5}$ concentrations, particularly in transportation-heavy metropolitan regions[25,26]. We leverage this period (March–April 2020) as a natural experiment that partially disentangles the confounding underlying legacy of historical social and economic policy from average air pollution exposures by providing a comparison between shutdown and non-shutdown (status quo) pollution distributions.

We employ established generalized difference-in-differences methods to quantify declines in ambient concentrations of two criteria pollutants, ground-level $PM_{2.5}$ and tropospheric column number density NO_2, during March–April 2020 and test for the existence of heterogeneous effects associated with the racial and ethnic composition of neighbourhoods. We utilize data from a relatively new network of low-cost particulate matter monitors that are predominantly privately owned and deployed outside homes, along with data from state-run air-quality sensors, satellite measurements, demographic and socioeconomic information, geographic data and cell-phone-based location data. By combining these datasets, we disentangle the contribution of local conditions (income, mobility, urban geography, weather) to local air pollution exposures. Data on mobility—defined as the extent to which individuals spend time away from their homes—are particularly important: they characterize variability in the shutdown's effect on the local activity of different communities, as essential worker status and economic insecurity are associated with less time spent at home[27,28].

Because the reduction of pollution when shutting down most of the in-person economy corresponds to the pollution burden created by that portion of the economy, pre-pandemic, we interpret statistically larger reductions in air pollution exposures for minority racial and ethnic groups—conditional on other confounding factors—as evidence of embedded bias in the generation and control of pollution from the in-person economy in the status quo. (We use the term 'in-person economy' to refer to economic activity from businesses, including transportation to and from those businesses, that closed during the COVID-19 economic shutdown. As outlined in Executive order N-33-20 of the State of California, closures affected all businesses except those in 16 critical infrastructure sectors: chemical; commercial facilities; communications; critical manufacturing; dams; defence industrial base; emergency service; energy; financial service; food and agriculture; government facility; healthcare and public health; information and technology; nuclear reactors, materials, and waste; transportation systems; and water and wastewater systems sectors. We note that while these sectors had the option to stay open, many did not, or continued operations in a reduced manner.) Our approach also demonstrates complementary inequities in the monitoring of pollution, shedding light on a path towards pro-actively addressing the identified inequities through air pollution monitoring policy that is itself environmentally just.

Results

Using daily and weekly pollution observations, along with demographic, geographic and mobility data, we estimate how much race and ethnicity alone explain the changes in air pollution exposures experienced during the COVID-19 shutdown in California. We account for time-varying factors (local mobility, weather, seasonality) and relatively static factors (population density, income, proximity to roads) known to contribute to heterogeneous pollution exposures. Our approach is described in detail in Methods as well as schematically in Fig. 1g and Extended Data Fig. 1.

Our study area and data are summarized in Fig. 1. Daily aerosol $PM_{2.5}$ measurements are drawn from a network of 830 monitors (106 public monitors from the California Air Resources Board (CARB) and 724 privately owned PurpleAir monitors; Extended Data Fig. 2a) and cover the period from 1 January to 30 April for both 2019 and 2020 to facilitate comparison across economic conditions at the same time of year. The low-cost PurpleAir sensors have been shown to correlate well with research-grade mass-based sensors, although they tend to have a high bias, which we have corrected before analysis (Methods and Supplementary Information). The $PM_{2.5}$ monitors are located in 733 unique census block groups (CBGs) across California. Satellite-derived tropospheric NO_2 (Fig. 1b) measurements from the TROPOMI instrument cover close to all 23,212 CBGs of California (22,503) but at an ~weekly timescale due to the overpass frequency of the Sentinel-5 precursor satellite. Local social, demographic and geographic characteristics (Fig. 1c–f; Methods), including income and population shares for race and Hispanic ethnicity, are heterogeneously distributed across the state; for example, income tends to be higher in coastal communities and cities, and the southeast and Central Valley regions have higher Hispanic population shares. Asian population share is highest in Los Angeles and the San Francisco Bay Area; Asian persons represent over 50% of the population in more than 1,400 CBGs. By contrast, fewer than 300 CBGs have majority Black populations, and these are spread more evenly throughout the state. (We use terminology derived from the census (Methods); Hispanic ethnicity is tallied independently of other race information in the United States and is therefore not mutually exclusive from race (Supplementary Fig. 1)). This complex human geography demonstrates the importance of rich measurement networks in addressing questions of environmental justice. The PurpleAir monitors provide a sevenfold increase in the number of sampled CBGs, although this increase still represents only 3.2% of all California block groups (Supplementary Information).

In 2019 (the year before the pandemic), without controlling for other sources of heterogeneity, areas with lower income and larger Black and Hispanic population shares were exposed to higher-than-average concentrations of both $PM_{2.5}$ and NO_2 compared with wealthier and White, non-Hispanic communities (Extended Data Fig. 3). Such descriptive air-quality differences have long been noted by environmental justice scholars and advocates[29–33], but these relationships are confounded by other drivers of pollution exposures (Extended Data Fig. 1). It is thus difficult to isolate any potential racial or ethnic bias in environmental policy from economic forces or other policies that contributed to the distributions of different populations around the state (for example, redlining)[34].

The COVID-19 pandemic temporarily removed a large portion of this confounding economic geography by 'turning off' most local in-person economic activity in the state. Figure 2 shows the depth and dimensions of this natural experiment across the state. The unique response to the spread of COVID-19, including stay-at-home orders, precipitated a steep decline in the average fraction of the day that people spent away from their homes (hereafter, mobility), which took a little under two weeks after the statewide emergency declaration (4 March 2020) to fully emerge (Fig. 2a). Importantly, reductions in time away from home did not occur equally for all state residents. CBGs with relatively high Hispanic population shares had both higher baseline mobility and much smaller mobility reductions during the shutdown than those with relatively low Hispanic population shares (Fig. 2b and Supplementary Table 1). This is probably due to the greater designation of essential jobs and economic vulnerability among Hispanic populations, relative to non-Hispanic populations, that preclude working from home[35]. This disparity is present, although much less pronounced, for CBGs with high and low Black population shares, and the pre- and post-shutdown differences are reversed for CBGs with high and low Asian population shares. We account for these different local responses in the statistical framework described below.

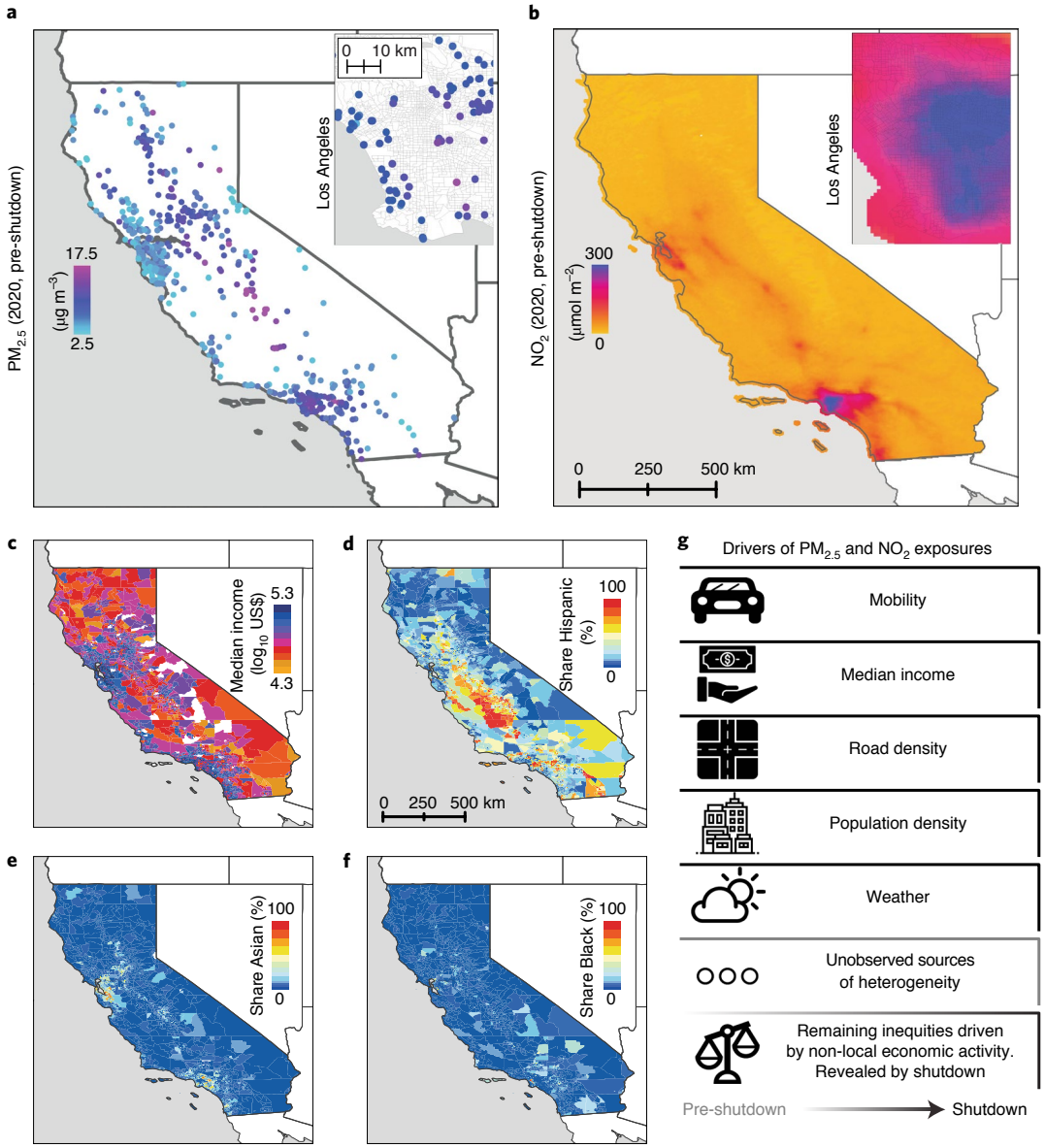


Fig. 1 | Pollution and demographic data used in this study. a,b, Average surface $PM_{2.5}$ (**a**) and tropospheric NO_2 (**b**) concentrations in the pre-shutdown period of 2020 in California, United States. **c**, Median household income (US$) in each CBG from the US Census Bureau 2018 5 yr American Community Survey (ACS). **d**–**f**, Share of the population in each CBG that is Hispanic (**d**), Asian (**e**) or Black (**f**), from the ACS. **g**, Schematic showing both slower-changing (assumed to be static over shorter periods) and higher-frequency factors that contribute to heterogeneous pollution exposures. Symbols in **g** courtesy of Noun Project: Automobile by Symbolon; Income and Highway by Vectors Point; Urban by Eucalup; weather by asianson.design; List by Richard Kunák; inequality by b farias.

Using a series of generalized difference-in-differences models (Methods), we estimate the relative magnitudes of the reductions in $PM_{2.5}$ and NO_2 concentrations before and after the shutdown (adjusting for 2019 concentrations) across different demographic gradients (we show and discuss $PM_{2.5}$ results in the main text, with NO_2 results in the Supplementary Information, for brevity). The best-fit coefficients for these models (Supplementary Tables 2, 3, 6 and 7, Fig. 3 and Extended Data Fig. 4) correspond to the statistically identifiable expected changes in air pollution, across the COVID-19 shutdown window, for a 0% versus 100% share of a given demographic group at the CBG level, or roughly a doubling of non-share variables (for example, income, road density, population density). (We note that estimating the population average change would require a stronger statistical assumption than we make about the similarity of other conditions between 2019 and 2020 (for example, seasonality; see Methods).) These coefficients show that lower-income neighbourhoods in California experienced greater reductions in $PM_{2.5}$ concentration (Fig. 3a); the positive and statistically significant coefficient for income indicates that lower incomes were strongly associated with a greater reduction of pollutant levels during shutdown. For example, our estimates indicate that a block group with an average income that is half that of a wealthier block group would have experienced a 1.0 μg m^{-3} greater reduction in $PM_{2.5}$ exposures. Changes in mobility, road density and population density at the level of a CBG are only weakly associated with changes in $PM_{2.5}$ concentrations (Fig. 3a).

We consider mobility to be a proxy for *l*ocal pollution-causing economic activity and assume that decreased mobility directly corresponded to reduced vehicle emissions along with a suite of local business-related emissions (for example, restaurant closures).

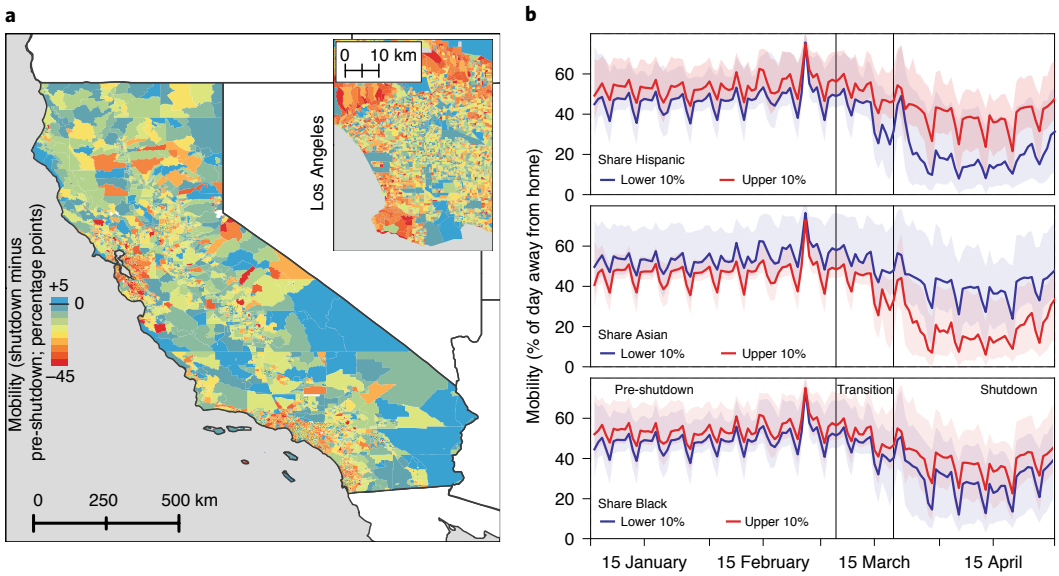


Fig. 2 | The COVID-19 'mobility shock'. a, Percentage-point difference in time spent at home pre-shutdown and during the shutdown at the CBG level in California, with an inset for the Los Angeles region. **b**, Mobility reductions for the median of the upper and lower 10% of three different population subsets. Shading indicates the 25th and 75th percentiles within each group. Vertical lines indicate the beginning and end of the transition (4 March 2020 and 19 March 2020) period excluded from our dynamic analysis. Average pre-/post-shutdown percentages are given in Supplementary Table 1.

Therefore, the relationship between the relative decline in local mobility and the relative decline in local air pollution gives insight into the pollution impacts of a block group's own economic activity. Figure 3b and Supplementary Table 4 show that residents in lower-income neighbourhoods reduced their mobility less than those in richer neighbourhoods during the shutdown period. Combined with the fact that lower-income areas experienced a larger drop in $PM_{2.5}$ concentrations, this finding suggests that local activity is not the primary driver of disparate exposures across the income gradient in California.

To further probe potential heterogeneity in the magnitude of shutdown impacts, we examine exposure changes across neighbourhood demographic gradients, with and without accounting for various local characteristics (Fig. 3c,e and Supplementary Tables 2 and 3). We identify substantial racial and ethnic disparities in air-quality improvements, even when accounting for income, road and population density and very fine-grained differences in weather patterns over space and time that strongly affect surface pollutant concentrations (Methods and Supplementary Information). We first examine the gradient for all non-White populations (that may also be Hispanic) and then decompose this group into the three largest racial and ethnic subgroups in the state. A ten-percentage-point increase in the non-White population share of a census block is associated with a 0.26 $\mu g\,m^{-3}$ reduction in $PM_{2.5}$ concentration after the shutdown. This falls to about 0.23 $\mu g\,m^{-3}$ once we include local mobility and allow for heterogeneous effects of the shutdown in terms of income, road density and population density (mobility impacts are shown directly in Fig. 3d and Supplementary Table 4). The decomposition reveals similar estimates for Hispanic and Asian populations, where we find reductions in $PM_{2.5}$ concentrations of 0.29 and 0.23 $\mu g\,m^{-3}$ per 0.1 increase in population share, respectively. We interpret this as evidence that in-person economic activity places a disproportionate pollution burden on non-White (and specifically Hispanic and Asian communities), only about a seventh of which is explained by differences in incomes and other location characteristics.

Hispanic and Asian are the two largest racial and ethnic minority groups in the state, making up about 39% and 16% of the population, respectively. While they share some similarities in historical inequitable treatment, there are several major differences in the socioeconomic attributes of the two groups. Acknowledging that we necessarily aggregate diverse subpopulations within racial and ethnic groups[36] (Supplementary Information), Asian Californians are predominately concentrated in urban areas and have on average higher incomes and education, whereas Hispanic populations are more skewed towards rural areas and have on average lower incomes and education.[37] Moreover, as described, the two groups had different baseline exposures and opposite average mobility responses to the shutdown. Despite these large circumstantial differences, their disproportionate exposure to economy-scale pollution is substantially similar, providing strong, albeit indirect, evidence of the influence of systemic racism in the mechanisms and institutions responsible for pollution control.

We do not find statistically significant pollution reductions associated with increased Black population share. That is, while the shutdown economy became more equitable in its pollution distribution vis-a-vis Hispanic and Asian communities, the same was not true for Black neighbourhoods, where substantial baseline pollution gradients remained unchanged. This may be in part statistical—the overall Black population share is around 7% and there are relatively few majority Black CBGs in the state (Fig. 1f)—but it also suggests that the in-person economy is not the main driver of pollution disparities for Black communities in California.

The findings from surface air-quality data are largely consistent with results from an analogous set of models using weekly satellite-derived NO_2 concentrations as the outcome, although some small differences between the two reflect both coverage discrepancies and the distributions of $PM_{2.5}$ and NO_2 sources in the state (Extended Data Fig. 4 and Supplementary Tables 6 and 7). The $PM_{2.5}$ results are also robust to consideration of sub-regions of California, for example, excluding Los Angeles, the Central Valley or both (Extended Data Fig. 5), suggesting that the findings are not driven by the seasonality of pollution, different pollution sources, demographics or unique airshed dynamics of these key regions. For both pollutants, the importance of accounting for fine-grained weather patterns is evident from the difference (for example, in Fig. 3

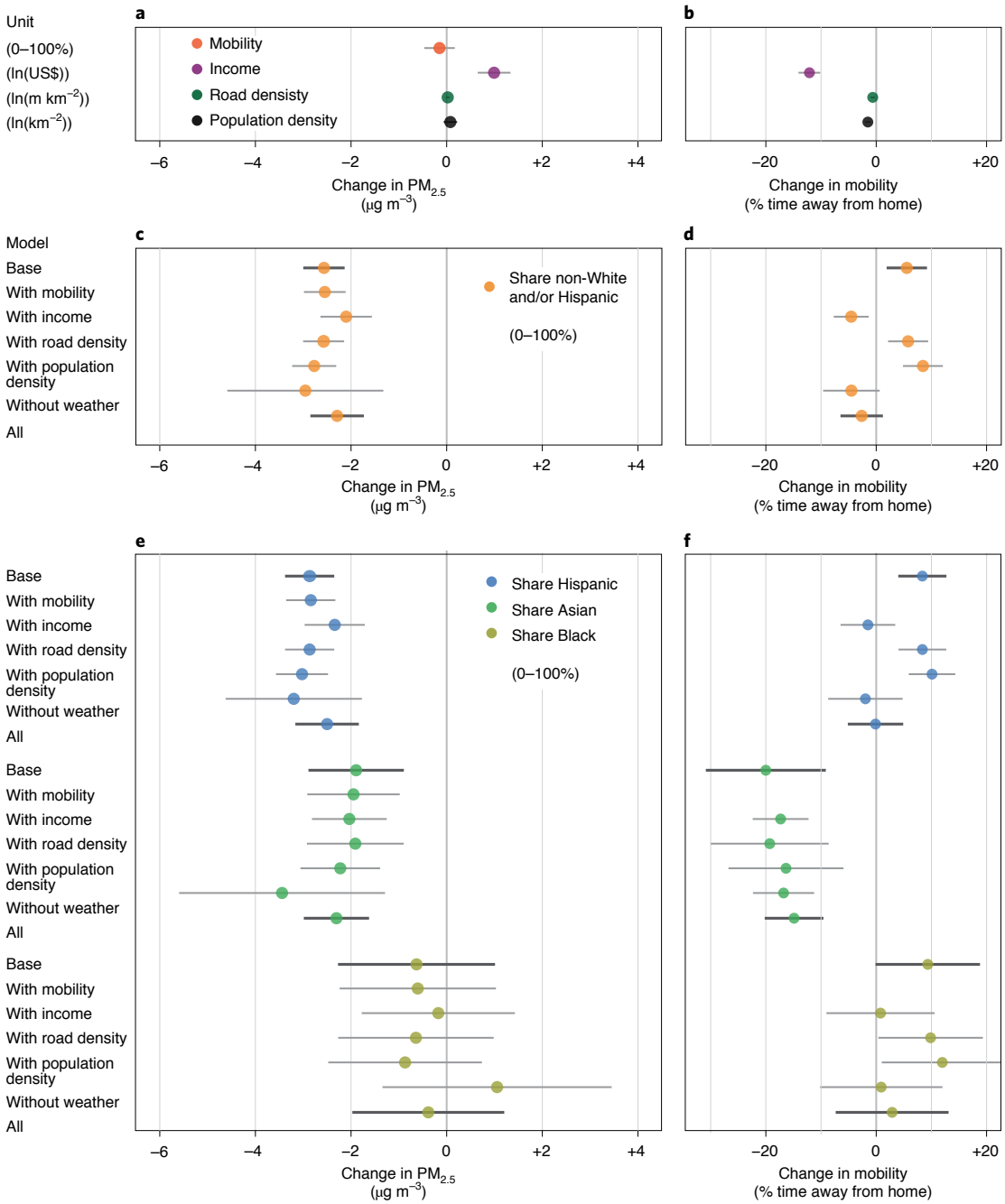


Fig. 3 | Impact of the economic shutdown on (left) $PM_{2.5}$ concentrations and (right) mobility. Points show heterogeneous changes across CBG characteristics estimated from difference-in-differences models, along with 95% confidence intervals. Intervals that include zero indicate that there was no differential reduction in exposures across the given gradient. **a**, Changes in daily $PM_{2.5}$ concentration across the shutdown estimated for various socioeconomic variables. The coefficient for mobility is the estimated difference between 0 and 100% of time spent at home; the coefficients for (ln) income, road and population density each represent the impact of an approximate doubling for each variable. **b**, Similar estimates with mobility as the dependent variable. **c**,**e**, Changes in $PM_{2.5}$ concentrations over the shutdown period across different racial and ethnic population shares, estimated with different physical and socioeconomic control variables (labels on left). The coefficients correspond to the expected changes between 0 and 100% population share at the CBG level. See Supplementary Tables 2 and 3 for values. **d**,**f**, Similar estimates, but with mobility as the outcome instead of $PM_{2.5}$. Values are given in Supplementary Tables 4 and 5. All four panels compare the post-shutdown difference from 2020 to 2019 to the pre-shutdown difference to account for seasonality. In addition, estimates were weighted to reflect the distribution of incomes, population shares and other location characteristics across all block groups in California and correct for the endogenous sampling of ground station locations (Extended Data Fig. 2b, Methods and Supplementary Information). The 'Base' model includes CBG and day-of-year fixed effects, as well as weather controls; other models incorporate the noted controls, or exclude weather, and 'All' includes everything.

and Extended Data Fig. 4) between estimates from our full model (All) and a model that includes all controls except weather ('w/o weather'). Prevailing weather patterns (Methods) that potentially transport pollution do account for some of the exposure disparities observed, with some variation by region and pollutant, but do not fully explain the observed patterns.

Discussion

In this Article, we provide new causal estimates of unequal air pollution exposure reductions experienced during California's COVID-19 economic shutdown. Because these reduction disparities are associated with widespread economic curtailment, they point to systemic racial and ethnic bias in the status quo generation and control of air pollution from the state's in-person economy. While this finding is robust to various specifications and data subsets, and is consistent across surface- and satellite-based data, our analysis nevertheless requires some contextualization and care in interpretation.

Importantly, while we note that exposure disparities are not explained by local mobility, the ability to fully distinguish local and non-local economic activity is limited, and representative spatial scales of local versus non-local may vary (for example, geographically, culturally or seasonally). We additionally note that while we focus on the non-local drivers of exposure disparities, local mobility-related pollution generation may nevertheless be caused by structurally unjust policy in other sectors (for example, housing or transportation). More broadly, it is important to recognize that contemporary and historical biases in other policy areas can lead to disparate average exposures, even if environmental policy surrounding the in-person economy were unbiased. This may be what explains our finding of higher average pollution exposures, but no disproportionate air-quality benefit from the COVID-19 shutdowns, for Black communities in California.

Our analysis consistently identified that lower-income communities in the state are disproportionately affected by pollution from the in-person economy. While we primarily employed income as a control, this income disparity represents an important concern in and of itself and presents policy challenges that are unique from those associated with combating institutional racism. California has one of the highest rankings for income inequality among US states[37], and our findings provide additional evidence that wealthier communities both are able to buy environmental quality (for example, via higher housing prices that embed air quality) and can afford to stay at home more fully during a pandemic.

Our empirical results complement a growing body of literature that uses chemical transport models (or reduced complexity models) to estimate pollutant exposures and map them to local socioeconomic and demographic characteristics. Importantly, such studies have been instrumental in identifying that a vast array of pollution sources contribute to baseline exposure disparities[15]. While these tools have become ever more powerful and accurate, and have the benefit of full coverage (compared with sparse monitoring networks), they do require accurate emissions inventories as inputs. Such inventories have rapidly improved in temporal resolution for long-lived greenhouse gases (for example, ref. [21]) but remain notoriously uncertain for air pollution, especially over short timescales and under abnormal economic conditions[38]. Our analysis strategy based on high-frequency observations of actual ground-level (or atmospheric column) pollution does not require emissions inventories and is thus well suited for understanding short-run changes in a way that modelling studies would be unable to capture.

Nevertheless, statistical studies like ours require accurate and unbiased characterization of the system under study and have their own shortcomings. For example, we cannot illuminate some of the more specific mechanisms of shutdown-induced variation in air pollution, such as whether being near to and downwind from a major road is more likely for ethnic/racial minorities[39,40]. A key point of contrast to modelling studies is that we do not explicitly account for individual point source emissions or wind, but instead use areal road density summaries (Methods), and detailed temperature, precipitation and relative humidity controls (see Extended Data Fig. 6), to capture much of this variation. Because our analysis focuses on differences within census blocks over time, the average influences of these and other unobserved factors are taken into account. Still, we cannot rule out that some of our measured effects may be driven by either variations in emissions or meteorological conditions that are correlated with both the demographic characteristics of a neighbourhood and the COVID-related shutdown. Future studies could focus on more thoroughly accounting for natural seasonal swings in air pollution and the full range of its spatial and temporal variability through the inclusion of more years of data[17]. This was not possible here due to the short timescale of PurpleAir and Sentinel data availability; however, our use of 2019 as a comparison for 2020, and the similarity of estimates made with pre- and post-shutdown 2020 data alone (Supplementary Table 8), underscore that the exposure disparities we estimate are not likely to be systematically changed by inclusion of more years of observations.

Beyond revealing disparities in pollution exposure generated by the in-person economy, our analysis also highlights inequality in local air pollution information. As we show, monitor placement matters for detection of exposure gradients. CARB recently re-focused air-quality monitoring in designated environmental justice communities[41], which has resulted in a more accurate sample of the state's Hispanic population distribution than (for example) PurpleAir. The PurpleAir monitoring network, established through the individual purchase and placement of (relatively) low-cost sensors, shows that citizen-science networks can be exceedingly useful for increasing the amount of public data but that those networks are unlikely to be optimally placed for addressing environmental justice questions (for which sensors are needed that accurately reflect the spatial distribution of all subpopulations). PurpleAir sensors also require care in correcting biases compared with monitoring-grade instruments[42]. On the public monitoring side, local governments that are responsible for choosing locations of sensors mandated by the Clean Air Act (that is, CARB) may also strategically place sensors to improve their chances of being in attainment[43,44]. This strategic placement reduces the ability of those sensor networks to detect environmental injustice[45] and makes adjustments for sampling bias, such as those proposed here, relevant for the larger literature (Supplementary Information). In addition, while we show that satellite-based observations can be helpful in understanding the spatial distribution of pollutants that underlies ground-based monitoring network samples, satellite data are spatially coarse compared with the average CBG and are more limited temporally (Supplementary Information). As such, satellites may not be able to replace ground-based monitoring when high spatial and temporal resolution are required. While a more spatially dense ground measurement network would vastly improve the ability to detect and address environmental injustice, reliability, cost, distribution and data curation would need to be considered in choosing a scale-up strategy[46,47].

Finally, while our analysis documents that the generation and control of pollution from California's in-person economy disproportionately and negatively affects the state's largest racial and ethnic minority communities, it also has potential applications in environmental policymaking. The United States has a multidecadal history of justifying environmental regulation through the use of an efficiency-based net benefit criterion (the simple objective that regulatory benefits exceed costs)[48,49]. Many state and federal entities additionally mandate that regulatory impact analysis include assessment of impacts to disadvantaged and vulnerable groups (for example, refs. [49,50]), and recently this has been identified as a major environmental policy priority[51]. It is difficult, however, for equity considerations to obtain equal footing with efficiency criteria when best practices for benefit–cost analysis are strictly codified[52], but there is no standard assessment criterion for justifying interventions that mitigate inequities[53]. For race- and ethnicity-based equity considerations, our methodology suggests that a net equalization criterion for environmental regulation could be constructed as follows and utilized in conjunction with benefit–cost analysis:

$$\text{Net Equalization} = \Delta\text{Exposure}_{\text{non-White}|\text{Income}} - \Delta\text{Exposure}_{\text{White,non-Hispanic}|\text{Income}} \quad (1)$$

In this framework, the pollution reduction effect we identify as associated with the COVID-19 shutdowns had a net equalizing effect (although absent inclusion of public health benefits, uniform reduction of regional economic activity would be unlikely to satisfy the net benefit criterion). (As in our analysis, controlling for income would be critical for any equalization criterion, not just because income can be a confounding factor in identifying environmental racism, but also because the policy mechanisms by which societies might address income-dependent environmental injustice are different from those for addressing environmental racism.) When systemic bias is driving adverse outcomes, public policy intervention focused narrowly on addressing market failures may eventually result in reversion to inequality[54]. Revising regulatory impact analysis protocols to include a clear, quantitative equalization standard would instigate a shift in focus of environmental regulation beyond efficient reduction of externalities.

Methods

$PM_{2.5}$ data. Surface station measurements of $PM_{2.5}$ were downloaded from publicly available PurpleAir and CARB databases (Extended Data Fig. 7). We downloaded all outdoor PurpleAir data available (1,891 individual stations) for January–April 2019 and 2020. PurpleAir sensors are relatively inexpensive and are usually privately owned, but much of the data is publicly available. The quality of these data are lower than that of data from regulatory monitors, but PurpleAir sensors provide unprecedented spatial coverage. Most PurpleAir sensors contain two Particulate Matter Sensor 5003 sensors (Plantower), which measure particle counts in six size bins. Counts are converted to $PM_{2.5}$ using two proprietary conversions, one intended for indoor use and the other for outdoor use; we use the outdoor conversion as recommended and tested by ref. [55]. We also average the two sensors (when available) and exclude days when daily $PM_{2.5}$ measurements within the same unit differ by at least 5 $\mu g\,m^{-3}$ and at least 16% (ref. [42]). In limited field evaluations, PurpleAir sensors have been shown to have strong correlations with high-quality sensors[55–58]. Ref. [55] also proposed a correction for effects of relative humidity, which we do not apply in part because we consider daily data rather than sub-daily. We do, however, apply a correction developed by the US Environmental Protection Agency, which tends to slightly over-correct the high bias of the PurpleAir instruments, meaning the presented results from these sensors are conservative (Supplementary Figure 2)[42].

We retrieved (1 May 2020) all hourly CARB $PM_{2.5}$ data in California available for January–April 2019 and 2020 using CARB's Air Quality and Meteorological Information System (150 individual stations). Professional instruments and oversight, particularly for calibration, provide higher confidence in the data quality of the CARB sites. However, there are an order of magnitude fewer CARB stations than PurpleAir sensors in California, which means studies using the government data are statistically limited by a relatively small sample size. Unlike PurpleAir sensors, CARB sites often offer a wide variety of air pollutant measurements, although we use only hourly $PM_{2.5}$ aggregated to the daily mean. For both CARB and PurpleAir data, days with mean $PM_{2.5}$ equal to zero or greater than 500 $\mu g\,m^{-3}$ are removed as outliers. Sites for which we remove more than 10% of data are excluded from the entire analysis. Sites with less than 80% data coverage during our study period are also excluded. For models that require 2019 and 2020 data, we apply these requirements to both years independently. This quality filtering removed 5.9% of daily CARB $PM_{2.5}$ data and 11.4% of daily PurpleAir data, resulting in data from 1,664 individual stations (119 CARB and 1,545 PurpleAir). However, only 830 of those (106 CARB and 724 PurpleAir) include data for 2019 and 2020 for the pre-shutdown and shutdown period and were therefore used in our empirical statistical analysis.

NO_2 data. We used the Copernicus Sentinel-5 Precursor Tropospheric Monitoring Instrument (TROPOMI, version 1.03.02) offline tropospheric NO_2 column number density[59] for mean NO_2 concentrations of the developed areas of each CBG. TROPOMI has a resolution of 0.01 arc degrees. Data were collected for January–April 2019 and 2020 and only for developed areas according to the US Geological Survey National Land Cover Database 2016[60]. For this study, all data were prepared using the Google Earth Engine Python API[61] and formatted as weekly means for each CBG. Weekly means were chosen to counteract the high frequency of missing data, particularly in northern California (Extended Data Fig. 8).

Climate data. For temperature, precipitation and relative humidity, we relied on the Gridded Surface Meteorological dataset[62]. This dataset provides daily information at 4 km resolution across the continental United States. For this study, data were aggregated in Google Earth Engine[61] in its original daily frequency (for $PM_{2.5}$ analysis) and as a weekly mean (for NO_2 analysis) for each CBG. The weekly mean data were aggregated only for developed areas according to the US Geological Survey National Land Cover Database 2016[60].

Mobility data. We use SafeGraph's Social Distancing Metrics[63], which were made available for research as part of the company's COVID-19 response and have been validated elsewhere (for example, ref. [64]). SafeGraph collects and cleans GPS pings from about 45 million mobile devices. The data are available daily at CBG resolution and are close to a random sample of the population. Our primary measure of mobility is not social distancing but rather the percentage of time spent away from home. We calculate this measure on the basis of the median time (in minutes) that a device was observed at its geohash-7 (about 153 m × 153 m) home location, which SafeGraph determines as the night-time residence of the device in the six weeks before. The data cover the entire period of observation from 1 January 2019 until the end of April 2020.

Demographic data. We downloaded CBG-level demographic information from the US Census Bureau 2018 5 yr ACS for all CBGs in California using the tidycensus package[65] for the R programming environment[66] (29 June 2020). Demographic features included ACS sample-based CBG-level estimates of population count; White race count (alone or in combination with one or more other races), or 'White'; Black or African American race count (alone or in combination with one or more other races), or 'Black'; Asian race count (alone or in combination with one or more other races), or 'Asian'; Hispanic or Latino origin (of any race) count, or 'Hispanic'; and median income. The other census race designations (American Indian or Alaska Native; Native Hawaiian or Other Pacific Islander) represent a substantially lower share of the California population and were therefore excluded from our analysis due to small sample sizes. The CBG-level 'share' of these groups was calculated by dividing the CBG count by the CBG population. Population density was calculated as the CBG population divided by the area of the CBG. For the aggregate comparison, we compute the share of the non-White population that may be Hispanic as one minus the share of Whites that do not also identify as Hispanic. Because Hispanic is a separate designation from race in the ACS (that is, those categorized as Hispanic may also be of any race), we evaluated how distinct Hispanic was from race variables of interest (Supplementary Fig. 1). On average, less than 1% of those identified at the CBG level as Hispanic were also identified as Black or Asian; 61% of Hispanic were White. Thus, Hispanic is effectively distinct from Asian and Black categorizations, and we consider Hispanic, Asian and Black designations to be unique demographic indicators in our model. The baseline reference group in the more detailed comparison contains all other races and ethnicities and therefore consists almost entirely of people who identify as non-Hispanic White.

Geographic data. We calculated road density ($m\,km^{-2}$) using The Global Roads Inventory Project (GRIP4)[67] vector dataset for North America. The GRIP4 dataset harmonizes global geospatial datasets on road infrastructure, including road features that can be categorized as highways, primary roads, secondary roads, tertiary roads and local roads. It is consistent with primary and secondary road classifications from the US Census TIGER/Line shapefiles for roads. To calculate road density for each CBG, we summed road lengths within the area of the CBG and divided by the area of the CBG. Calculations were done using the sf package[68] in the R programming environment[66].

Study period and design. We consider three periods between 1 January and 30 April in 2019 and 2020. The first period is 'pre-shutdown,' followed by a 'transition' and then 'shutdown.' The transition is defined as the period between the statewide emergency declaration (4 March 2020) and the statewide stay-at-home order (19 March 2020). The mobility data demonstrate that activity declined throughout this period (Fig. 2). This is consistent with recent literature that shows that fear was a potent driver of the decline in mobility and often pre-empted county-wide legal restrictions[69]. The shutdown period begins with the stay-at-home order and ends at the end of our study period. We exclude the transition from the analyses described below. This precludes the use of variation in treatment timing to assist with causal identification. Instead, we proceed by using the interaction between the shutdown and racial or ethnic composition of CBGs as the treatment, allowing us to directly estimate the additional pollution burden of economic activities that were halted during the shutdown on block groups with certain demographic compositions. This is standard practice, referred to as 'generalized difference in differences' (see Supplementary Information for details). In our case, there is no group that remains untreated and no variation in treatment timing but heterogeneity in treatment intensity.

Empirical strategy. In our statistical analyses, our main dependent variable is an (average) measure of air quality ($PM_{2.5}$ or NO_2) in CBG i at day (or week) t. We focus on block groups to minimize the influence of aggregation bias or the 'ecological fallacy'[70] and study temporal variation in air quality across block groups using a difference-in-differences design. Difference-in-differences methods are commonly used to study causal effects in economics[71]. Our objective is to estimate

the heterogeneity in the effect of the shutdown across different communities rather than the overall effect of the shutdown. We focus on the racial and ethnic composition of CBGs using population shares of California's three biggest racial and ethnic minority groups (Hispanic, Asian and Black). We first establish the existence of air pollution exposure reduction inequities and then include a rich set of controls to assess the racial and ethnic inequities that remain after accounting for differences in mobility, income and location[72–74].

A key concern is that differences in air quality are driven by interannual cycles in pollution and particle concentration that are unrelated to the shutdown[75]. We address this issue in several ways. First, we subtract observed air quality in 2019 from the 2020 value. All annual differences, after aligning the weekdays, are denoted by $\tilde{y}_{it} = y_{it} - y_{i,t-364}$. Second, we flexibly control for local weather conditions in 2020 and 2019 (see Extended Data Fig. 6 and Supplementary Information for details on selection of weather controls). Finally, we allow for a rich set of day or week fixed effects that capture the remaining differences in synoptic scale weather patterns. Putting all this together, we use Ordinary Least Squares to estimate the heterogeneous effect of the shutdown using variants of the following specification:

$$\tilde{y}_{it} = \sum_k \gamma_k \left(d_t \times x_i^k \right) + \theta \tilde{M}_{it} + f^{2020}(T, \mathrm{RH}, P)_{it} + f^{2019}(T, \mathrm{RH}, P)_{it} + \lambda_t + \mu_i + e_{it} \quad (2)$$

where d_t is an indicator for the post-shutdown period, x_i^k are the population shares of the three minority groups studied here or other (relatively) time-invariant location characteristics k that vary at the CBG level, $\tilde{M}_{it}$ is the annual difference in mobility on day t in CBG i (constructed analogously to $\tilde{y}_{it}$), $f^{2020}(\cdot)_{it}$ and $f^{2019}(\cdot)_{it}$ approximate the nonlinear response of pollution and particle concentration to weather with interacted fixed effects for each decile of temperature (T), relative humidity (RH) and precipitation (P) in the corresponding year, λ_t are day (or week) fixed effects, μ_i are CBG fixed effects (capturing changes in the number of stations in a block group across years), and e_{it} is an error term. We cluster all standard errors on the county level, as stay-at-home and local health ordinances are spatially and temporally correlated at this level.

We are interested in γ_k, which captures the heterogeneous impact of the shutdown across different demographic gradients (see Supplementary Information for a derivation). The baseline effect of the shutdown, d_t, is not statistically identified without the assumption of constant seasonal emissions patterns, as that baseline effect occurs simultaneously for all block groups in California (Extended Data Fig. 9) and is therefore collinear with seasonal shifts in air quality that are unrelated to the COVID-19 shutdowns. Heterogeneous impacts are identified by variation among block groups experiencing a COVID-19 shutdown-related air pollution change only and can be interpreted as the effect of the shutdown relative to some baseline. This requires a weaker assumption: that the interannual differences in pollution are not simultaneously correlated with the timing of the shutdown and the spatial distribution of race and income. Our weather controls make this a plausible assumption by accounting for systematic differences in temperature, humidity and rainfall across different parts of the state. We interpret the coefficient on $d_t \times$ %Hispanic as the difference in pollutant concentration for a block group that is 100% Hispanic relative to a block group that is 0% Hispanic. Differences in air pollution concentrations across the shutdown window are typically reductions in air pollution, which we consider to be equivalent to the expected increase after a return to 'business-as-usual' conditions. We address the representativity of monitor placement in both CARB and PurpleAir networks by comparing estimates based on each network alone, both networks combined, and both combined with weights derived to make the sampled locations (that is, CBGs with monitors) match state demographics (see Extended Data Figs. 2b and 10, and Supplementary Information for details).

Software. All data processing and analysis other than acquisition and pre-processing of mobility information were done using the R programming environment[66] and the Python API for Google Earth Engine[61]. ArcGIS was used to make the maps in Fig. 1 and Extended Data Figs. 2 and 8, and Adobe Illustrator was used for final assembly of composite figures.

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

Data availability

The pollution, weather and demographic data used in this project are all publicly available. PurpleAir data are accessible via public API, CARB data can be found at https://www.arb.ca.gov/aqmis2/aqmis2.php. Gridded NO_2 data (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_NRTI_L3_NO2) and GridMet weather data (https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_GRIDMET) are available via Google Earth Engine. Mobility data from Safegraph are available to academic researchers (https://www.safegraph.com/academics) but may not be reposted in raw format. GRIP4 road data are available at https://www.globio.info/download-grip-dataset, and ACS data were accessed via R using the tidycensus package. All publicly available raw data for this project can be found in our project dataverse (https://doi.org/10.7910/DVN/ZXVB7A), and processed and derived data can be found in our GitHub repository (https://github.com/jaburney/CA-COVIDEJ-2022). Finally, we have provided an online tool to facilitate exploration of the data at the CBG scale (https://sabenz.users.earthengine.app/view/covid-ej).

Code availability

Replication code is maintained in our GitHub repository (https://github.com/jaburney/CA-COVIDEJ-2022).

Received: 21 April 2021; Accepted: 26 January 2022;
Published online: 07 April 2022

References

1. Kendi, I. X. *How to Be an Antiracist* (One World, 2019).
2. Feagin, J. *Systemic Racism: A Theory of Oppression* (Routledge, 2013).
3. Krieger, N. et al. Structural racism, historical redlining, and risk of preterm birth in New York City, 2013–2017. *Am. J. Public Health* **110**, 1046–1053 (2020).
4. Pellow, D. N. Environmental inequality formation: toward a theory of environmental injustice. *Am. Behav. Sci.* **43**, 581–601 (2000).
5. Bullard, R. D. The legacy of American apartheid and environmental racism. *J. Civ. Rights Econ. Dev.* **9**, 3 (1993).
6. Banzhaf, S., Ma, L. & Timmins, C. Environmental justice: the economics of race, place, and pollution. *J. Econ. Perspect.* **33**, 185–208 (2019).
7. Hoek, G. et al. Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environ. Health* **12**, 43 (2013).
8. Burnett, R. et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl Acad. Sci. USA* **115**, 9592–9597 (2018).
9. Nardone, A. et al. Associations between historical residential redlining and current age-adjusted rates of emergency department visits due to asthma across eight cities in California: an ecological study. *Lancet Planet. Health* **4**, e24–e31 (2020).
10. Tessum, C. W. et al. Inequity in consumption of goods and services adds to racial–ethnic disparities in air pollution exposure. *Proc. Natl Acad. Sci. USA* **116**, 6001–6006 (2019).
11. Mikati, I., Benson, A. F., Luben, T. J., Sacks, J. D. & Richmond-Bryant, J. Disparities in distribution of particulate matter emission sources by race and poverty status. *Am. J. Public Health* **108**, 480–485 (2018).
12. *Environmental Justice* (United States Environmental Protection Agency, 2014); https://www.epa.gov/environmentaljustice
13. Seaman, N. L. Meteorological modeling for air-quality assessments. *Atmos. Environ.* **34**, 2231–2259 (2000).
14. O'Neill, M. S. et al. Health, wealth, and air pollution: advancing theory and methods. *Environ. Health Perspect.* **111**, 1861–1870 (2003).
15. Tessum, C. W. et al. $PM_{2.5}$ polluters disproportionately and systemically affect people of color in the United States. *Sci. Adv.* **7**, eabf4491 (2021).
16. Chay, K. Y. & Greenstone, M. The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *Q. J. Econ.* **118**, 1121–1167 (2003).
17. Diffenbaugh, N. S. et al. The COVID-19 lockdowns: a window into the Earth system. *Nat. Rev. Earth Environ.* **1**, 470–481 (2020).
18. *Stay Home Q&A* (State of California, 2020); https://covid19.ca.gov/stay-home-except-for-essential-needs/
19. LeQuéré, C. et al. Temporary reduction in daily global CO_2 emissions during the COVID-19 forced confinement. *Nat. Clim. Change* **10**, 647–653 (2020).
20. *Business Response Survey to the Coronavirus Pandemic* (US Bureau of Labor Statistics, 2020); www.bls.gov/brs/
21. Liu, Z. et al. Near-real-time monitoring of global CO_2 emissions reveals the effects of the COVID-19 pandemic. *Nat. Commun.* **11**, 5172 (2020).
22. Cai, J. et al. Application of land use regression to assess exposure and identify potential sources in $PM_{2.5}$, BC, NO_2 concentrations. *Atmos. Environ.* **223**, 117267 (2020).
23. Davis, M. E. Recessions and health: the impact of economic trends on air pollution in California. *Am. J. Public Health* **102**, 1951–1956 (2012).
24. Thakrar, S. K. et al. Reducing mortality from air pollution in the United States by targeting specific emission sources. *Environ. Sci. Technol. Lett.* **7**, 639–645 (2020).
25. Venter, Z. S., Aunan, K., Chowdhury, S. & Lelieveld, J. COVID-19 lockdowns cause global air pollution declines. *Proc. Natl Acad. Sci. USA* **117**, 18984–18990 (2020).
26. Liu, F. et al. Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19. *Sci. Adv.* **6**, eabc2992 (2020).
27. Weill, J. A., Stigler, M., Deschenes, O. & Springborn, M. R. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proc. Natl Acad. Sci. USA* **117**, 19658–19660 (2020).

28. Bonaccorsi, G. et al. Economic and social consequences of human mobility restrictions under COVID-19. *Proc. Natl Acad. Sci. USA* **117**, 15530–15535 (2020).
29. Perlin, S. A., Setzer, R. W., Creason, J. & Sexton, K. Distribution of industrial air emissions by income and race in the United States: an approach using the toxic release inventory. *Environ. Sci. Technol.* **29**, 69–80 (1995).
30. Gwynn, R. C. & Thurston, G. D. The burden of air pollution: impacts among racial minorities. *Environ. Health Perspect.* **109**, 501–506 (2001).
31. Pastor, M. Jr, Morello-Frosch, R. & Sadd, J. L. The air is always cleaner on the other side: race, space, and ambient air toxics exposures in California. *J. Urban Aff.* **27**, 127–148 (2005).
32. Nguyen, N. P. & Marshall, J. D. Impact, efficiency, inequality, and injustice of urban air pollution: variability by emission location. *Environ. Res. Lett.* **13**, 024002 (2018).
33. Clark, L. P., Millet, D. B. & Marshall, J. D. Changes in transportation-related air pollution exposures by race–ethnicity and socioeconomic status: outdoor nitrogen dioxide in the United States in 2000 and 2010. *Environ. Health Perspect.* **125**, 097012 (2017).
34. Nardone, A. et al. Associations between historical residential redlining and current age-adjusted rates of emergency department visits due to asthma across eight cities in California: an ecological study. *Lancet Planet. Health* **4**, e24–e31 (2020).
35. Galdamez, M., Kesteven, C. & Melaas, A. *In a Vulnerable State: Hispanic Essential Workers in California (*Milken Institute, 2020).
36. Lee, J., Ramakrishnan, K. & Wong, J. Accurately counting Asian Americans is a civil rights issue. *Ann. Am. Acad. Pol. Soc. Sci.* **677**, 191–202 (2018).
37. *ACS Provides New State and Local Income, Poverty and Health Insurance Statistics* (US Census Bureau, 2019).
38. Hoesly, R. M. et al. Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emission Data System (CEDS). *Geosci. Model Dev.* **11**, 369–408 (2018).
39. Anderson, M. L. As the wind blows: the effects of long-term exposure to air pollution on mortality. *J. Eur. Econ. Assoc.* **18**, 1886–1927 (2020).
40. Deryugina, T., Heutel, G., Miller, N. H., Molitor, D. & Reif, J. The mortality and medical costs of air pollution: evidence from changes in wind direction. *Am. Econ. Rev.* **109**, 4178–4219 (2019).
41. *AB-617: Nonvehicular Air Pollution: Criteria Air Pollutants and Toxic Air Contaminants* (California State Assembly, 2017).
42. Barkjohn, K., Gantt, B. & Clements, A. Development and application of a United States-wide correction for $PM_{2.5}$ data collected with the PurpleAir sensor. *Atmos. Meas. Tech.* **14**, 4617–4637 (2021).
43. Fowler, L. Local governments: the 'hidden partners' of air quality management. *State Local Gov. Rev.* **48**, 175–188 (2016).
44. Muller, N. Z. & Ruud, P. A. What forces dictate the design of pollution monitoring networks? *Environ. Model. Assess.* **23**, 1–14 (2018).
45. Grainger, C. & Schreiber, A. Discrimination in ambient air pollution monitoring? *AEA Pap. Proc.* **109**, 277–282 (2019).
46. Robinson, D. L. Accurate, low cost $PM_{2.5}$ measurements demonstrate the large spatial variation in wood smoke pollution in regional Australia and improve modeling and estimates of health costs. *Atmosphere* **11**, 856 (2020).
47. Becnel, T. et al. A distributed low-cost pollution monitoring platform. *IEEE Internet Things J.* **6**, 10738–10748 (2019).
48. Food and Drug Administration. Executive Order 12291: Federal Regulation. *Federal Register* **46** (1981); https://www.archives.gov/federal-register/codification/executive-order/12291.html
49. Food and Drug Administration. Executive Order 12866: Regulatory Planning and Review. *Federal Register* **58**, 190 (1993); https://www.archives.gov/files/federal-register/executive-orders/pdf/12866.pdf
50. Food and Drug Administration. Executive Order 12898: Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations. *Federal Register* **59**, 32 (1994); https://www.archives.gov/files/federal-register/executive-orders/pdf/12898.pdf
51. Food and Drug Administration. Executive Order 14008: Tackling the Climate Crisis at Home and Abroad. *Federal Register* **86**, 19 (2021); https://www.regulations.gov/document/EPA-HQ-OPPT-2021-0202-0012
52. *Circular A-4: Regulatory Analysis* (Office of Management and Budget, 2003).
53. Hahn, R. W. Equity in cost–benefit analysis. *Science* **372**, 439–439 (2021).
54. Yang, T. Melding civil rights and environmentalism: finding environmental justice's place in environmental regulation. *Harvard Environ. Law Rev.* **26**, 4–8 (2002).
55. Tryner, J. et al. Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers. *Atmos. Environ.* **220**, 117067 (2020).
56. Sayahi, T., Butterfield, A. & Kelly, K. E. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environ. Pollut.* **245**, 932–940 (2019).
57. Bi, J., Wildani, A., Chang, H. H. & Liu, Y. Incorporating low-cost sensor measurements into high-resolution $PM_{2.5}$ modeling at a large spatial scale. *Environ. Sci. Technol.* **54**, 2152–2162 (2020).
58. Stavroulas, I. et al. Field evaluation of low-cost PM sensors (Purple Air PA-ii) under variable urban air quality conditions, in Greece. *Atmosphere* **11**, 926 (2020).
59. Veefkind, J. P. et al. TROPOMI on the ESA Sentinel-5 precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens. Environ.* **120**, 70–83 (2012).
60. Yang, L. et al. A new generation of the United States National Land Cover database: requirements, research priorities, design, and implementation strategies. *ISPRS J. Photogramm. Remote Sens.* **146**, 108–123 (2018).
61. Gorelick, N. et al. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
62. Abatzoglou, J. T. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.* **33**, 121–131 (2011).
63. *Social Distancing Metrics* (SafeGraph, 2020).
64. Yan, Y. et al. Measuring voluntary and policy-induced social distancing behavior during the COVID-19 pandemic. *Proc. Natl Acad. Sci. USA* **118**, e2008814118 (2021).
65. Walker, K. tidycensus: Load US Census Boundary and Attribute Data as `tidyverse' and `sf'-Ready Data Frames. R package version 0.9.9.5 (2020).
66. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
67. Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J. & Schipper, A. M. Global patterns of current and future road infrastructure. *Environ. Res. Lett.* **13**, 064006 (2018).
68. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R J.* **10**, 439–446 (2018).
69. Goolsbee, A. & Syverson, C. *Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020* Working Paper 27432 (NBER, 2020).
70. Baden, B. M., Noonan, D. S. & Turaga, R. M. R. Scales of justice: is there a geographic bias in environmental equity analysis? *J. Environ. Plan. Manage.* **50**, 163–185 (2007).
71. Wooldridge, J. M. Asymptotic properties of weighted *M*-estimators for variable probability samples. *Econometrica* **67**, 1385–1406 (1999).
72. Banzhaf, S., Ma, L. & Timmins, C. Environmental justice: the economics of race, place, and pollution. *J. Econ. Perspect.* **33**, 185–208 (2019).
73. Bonaccorsi, G. et al. Economic and social consequences of human mobility restrictions under COVID-19. *Proc. Natl Acad. Sci. USA* **117**, 15530–15535 (2020).
74. Warren, M. S. & Skillman, S. W. Mobility changes in response to COVID-19. Preprint at *arXiv* https://arxiv.org/abs/2003.14228 (2020).
75. Hasheminassab, S. et al. Spatial and temporal variability of sources of ambient fine particulate matter ($PM_{2.5}$) in California. *Atmos. Chem. Phys.* **14**, 12085–12097 (2014).

Acknowledgements

The United States National Science Foundation NSF/USDA NIFA INFEWS T1 no. 1619318 supported J.A.B., M.C.L. and S.A.B.; NSF CNH-L no. 1715557 supported J.A.B. and P.P. The Alexander von Humboldt Foundation supported R.B.; the Stanford Woods Institute for the Environment supported K.S.H. The Big Pixel Initiative at UC San Diego and the Banting postdoctoral fellowship, administered by the Government of Canada, supported S.A.B.

Author contributions

K.S.H., P.P., J.A.B., M.C.L, S.A.B., L.C.S. and R.B. designed study. S.A.B., M.C.L., K.S.H., P.P. and R.B. prepared data and code for analyses; R.B. and L.C.S. ran statistical models. All authors interpreted results, prepared figures, wrote and edited the manuscript together.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41893-022-00856-1.

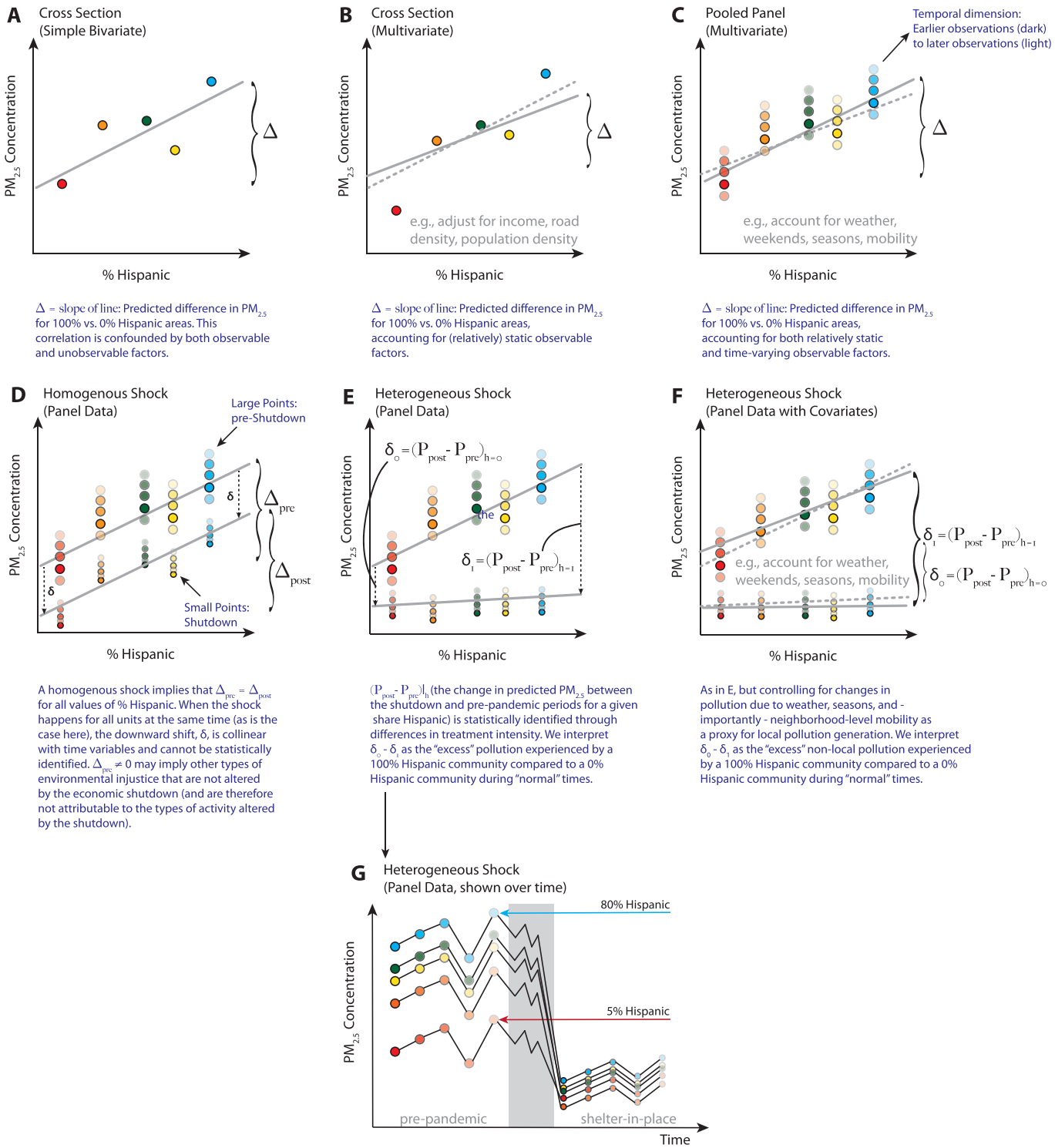
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41893-022-00856-1.

Correspondence and requests for materials should be addressed to Jennifer A. Burney.

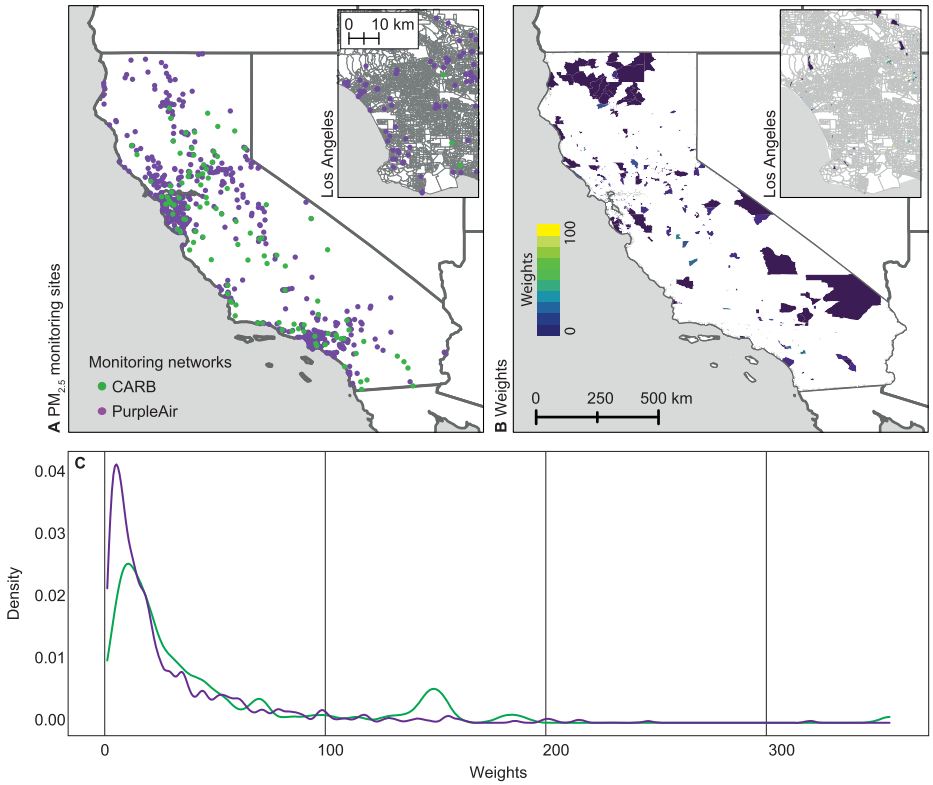
Peer review information *Nature Sustainability* thanks Dena Montague and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

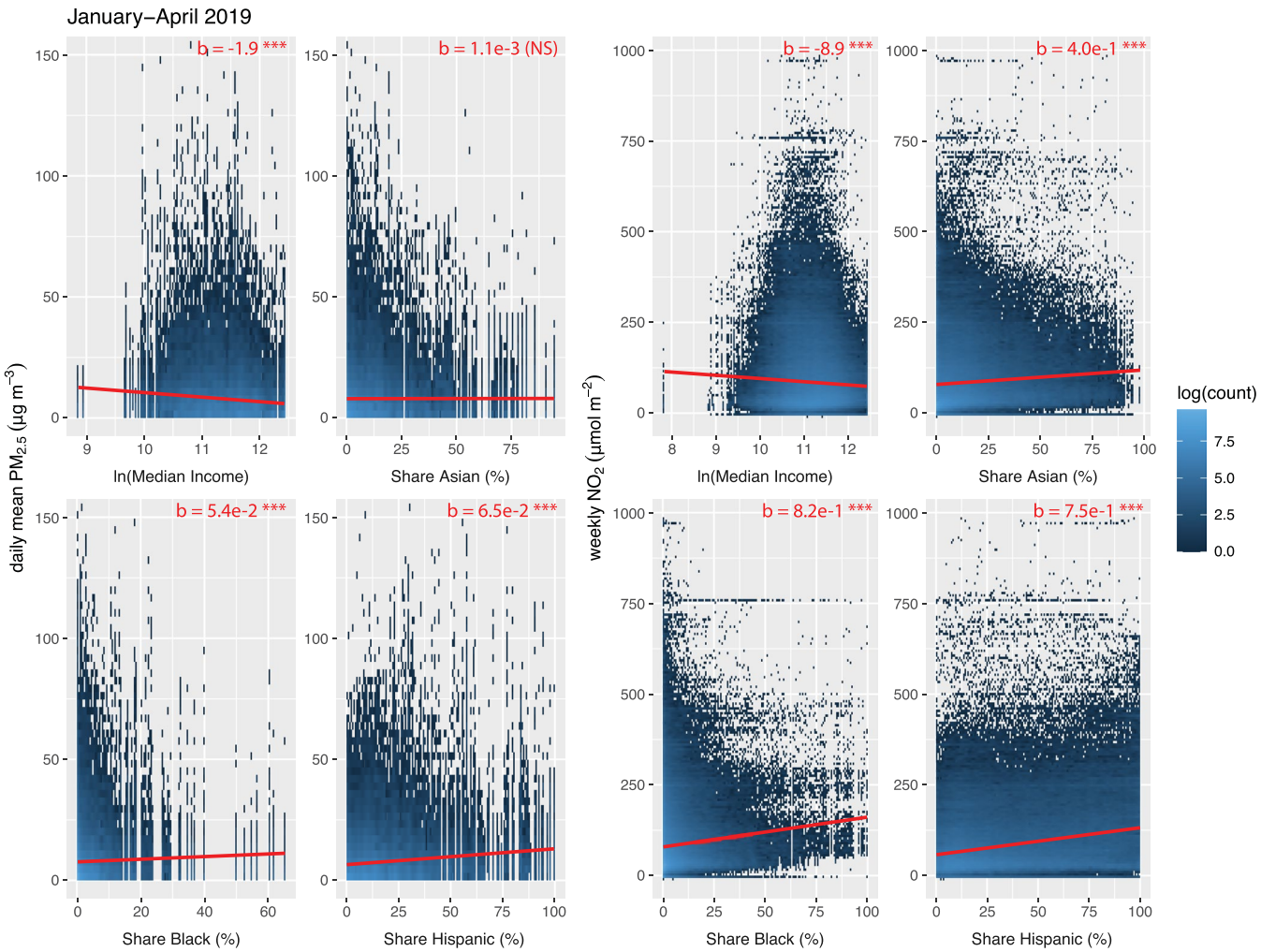
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



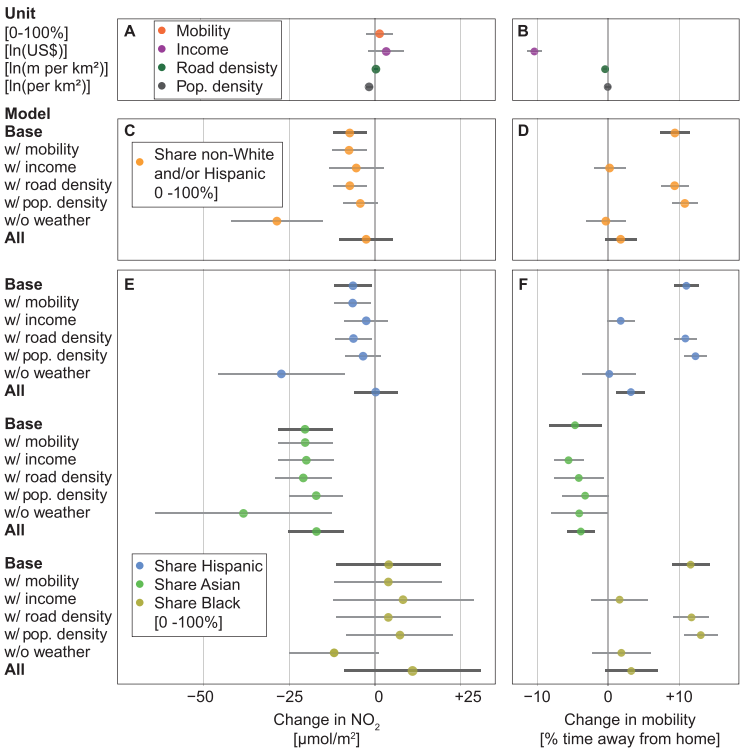
Extended Data Fig. 1 | Why the COVID-19 'shock' offers new insight into questions of environmental justice. For simplicity, imagine 5 communities across (for example) a state, represented by the five colors here. These locations each have a different racial/ethnic composition, represented here for simplicity in one dimension, as the share of the population that is Hispanic. Many observations of environmental injustice rely on cross-sectional analyses, either **A** without or **B** accounting for potential slower-moving confounds. **C** However, many high-frequency variables contribute to ambient pollution levels and might be correlated with geography and socioeconomic variables; panel analysis with repeat observations over time allows for inclusion of these types of covariates, and can thus account for the contributions that (for example) natural weather patterns make to exposure disparities. However, even a panel analysis is subject to potential confounding, and interpretation of residual exposure disparities as environmental injustice caused by the economy remains problematic. The COVID-19 economic shock creates a large perturbation that 'turns off' a portion of the economy, and thus reveals the footprint of pollution caused by that in-person economic activity. We test for whether this shock changes exposure gradients (that is, whether the shock looks like **D** or **E**, and as such whether the in-person economy is contributing to environmental racism. **F** The ability to account for mobility in this framework further allows the separation of very local activity from broader activity (see Supplementary Information Text). (For clarity, **G** shows the heterogeneous shock in time series).



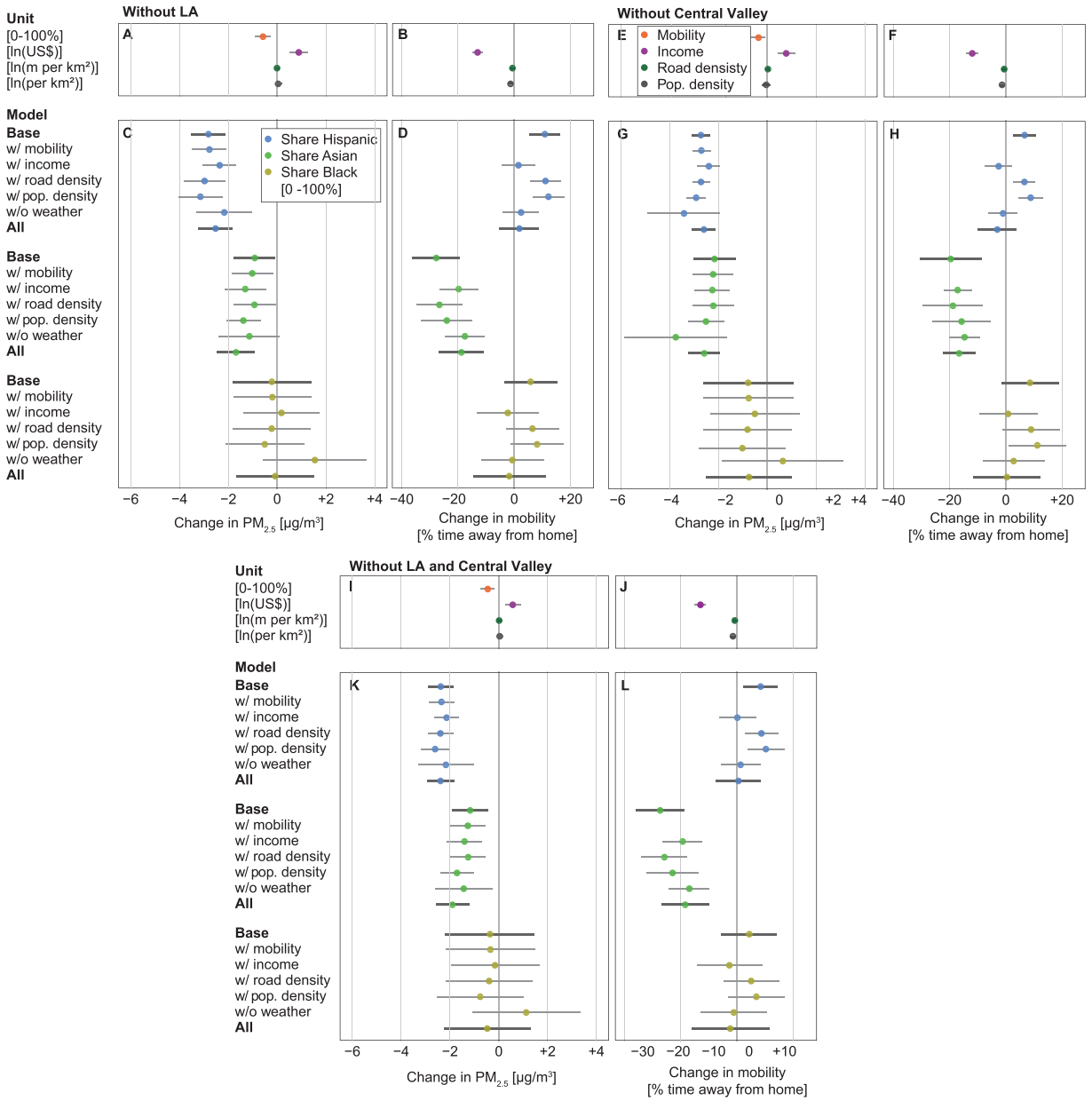
Extended Data Fig. 2 | $PM_{2.5}$ networks and weights. A Location of public (California Air Resources Board (CARB) and United States Environmental Protection Agency (EPA)) $PM_{2.5}$ monitors, as well as privately-owned PurpleAir $PM_{2.5}$ monitors used in this study. **B** Weights used in the model to better represent the Californian population. Each dot represents one census block group. **C** Distribution of weights generated by the raking process, across sensor types.



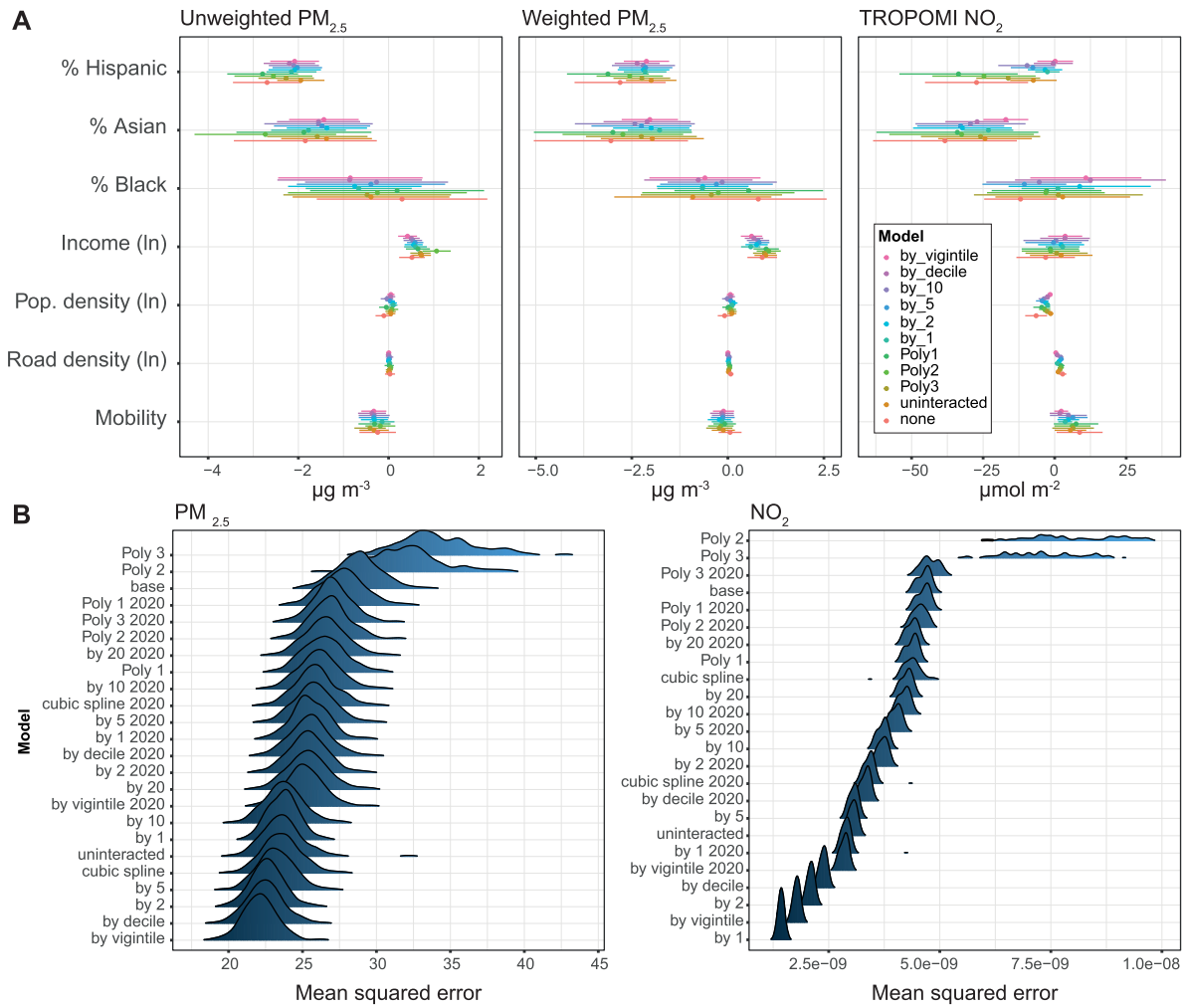
Extended Data Fig. 3 | Baseline associations between income, race, ethnicity and pollution exposure. Relationships between pooled January-April 2019 $PM_{2.5}$ and NO_2 observations and census block group demographics, without controlling for other sources of heterogeneity. The line represents the best linear fit. Red text shows slope of fit line, *** indicates p < 1e-3; NS = not statistically significant.



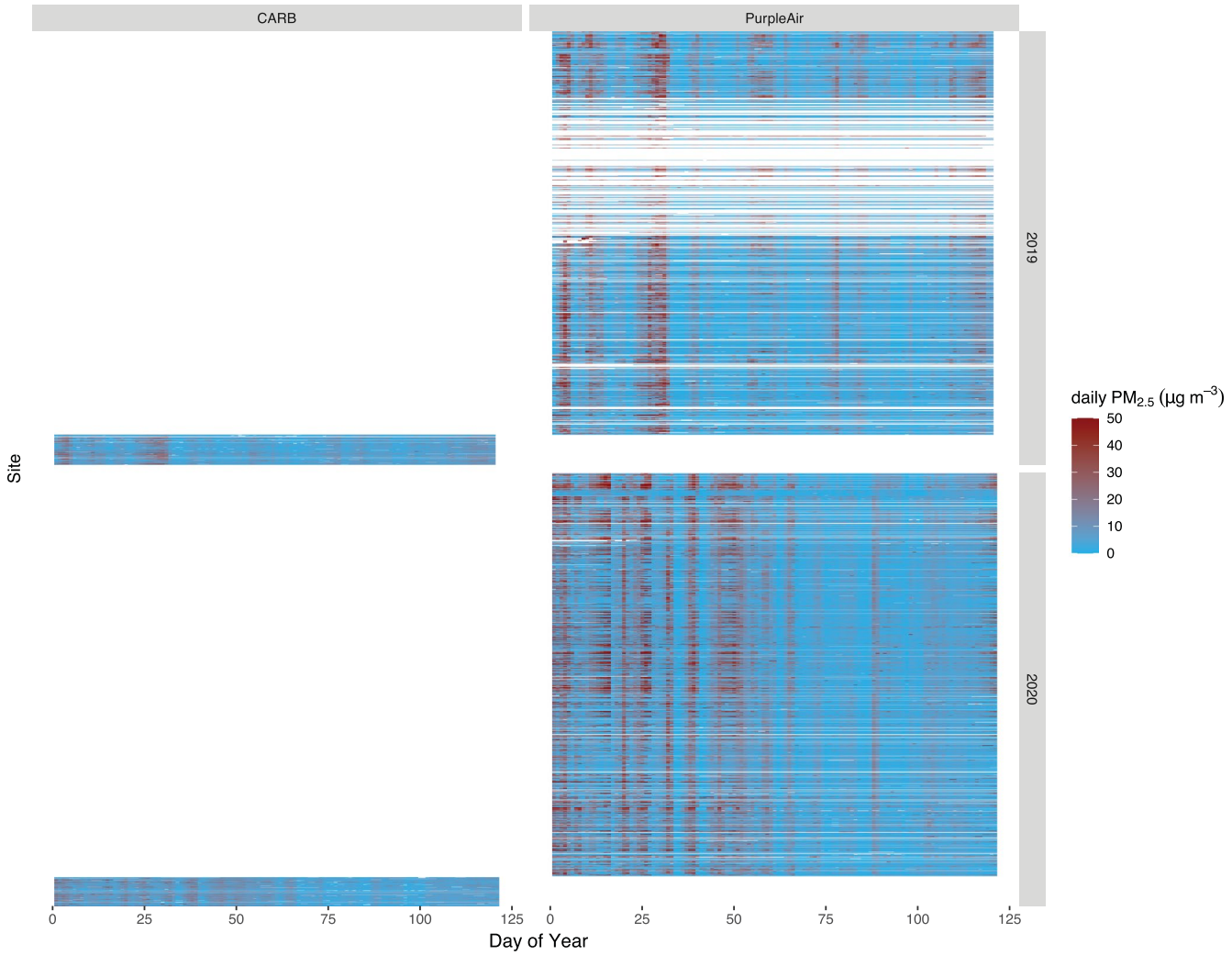
Extended Data Fig. 4 | Same as Fig. 3, but for weekly NO_2. Since NO_2 measurements are from the TROPOMI satellite instrument, they cover all census block groups in California. Qualitative differences between NO_2 estimates (A,C,E) and Fig. 3a,c,e thus represent a combination of coverage, and differences in $PM_{2.5}$ and NO_2 distributions. Differences in mobility between these estimates (B,D,F) and Fig. 3b,d,f are due to coverage differences (representing mobility in all census block groups here and mobility census block groups with Purple Air stations for Fig. 3) and weekly versus daily analysis. Values of the coefficients in parts A and C are given in Table S6 and part E in Table S7.



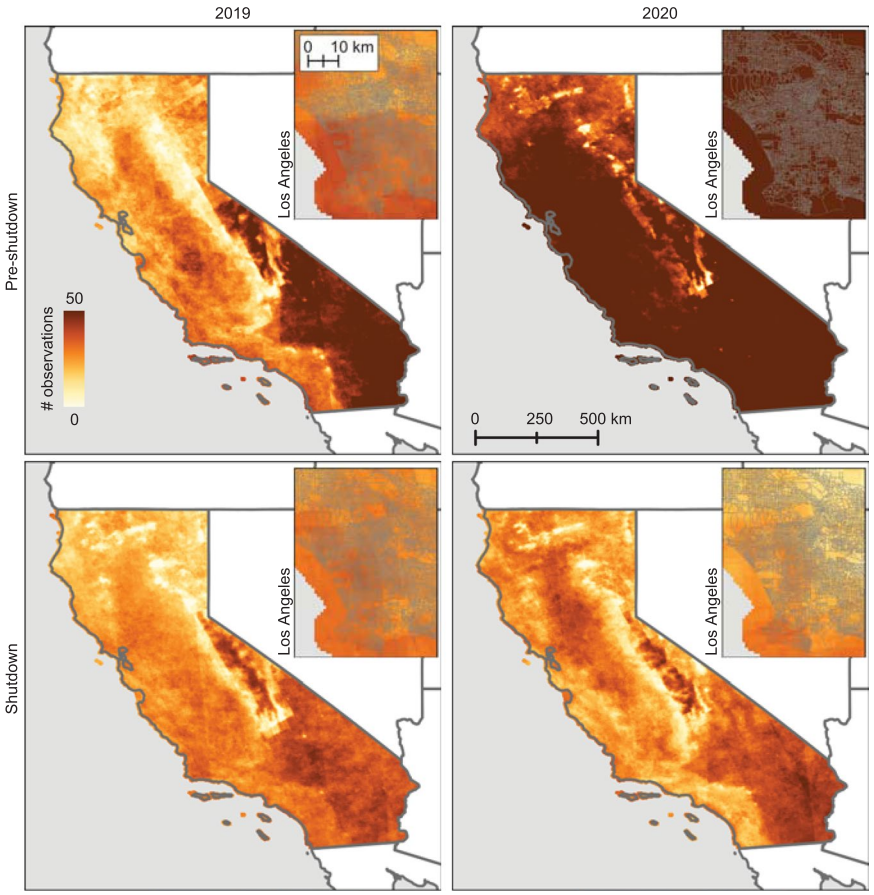
Extended Data Fig. 5 | Same as Fig. 3, but excluding Los Angeles County (A-D), the Central Valley (E-H) and both (I-L). There is no change in patterns when re-estimating our models without these regions, highlighting that climate or demographic differences in these key regions is not a key driver of the total observed differences.



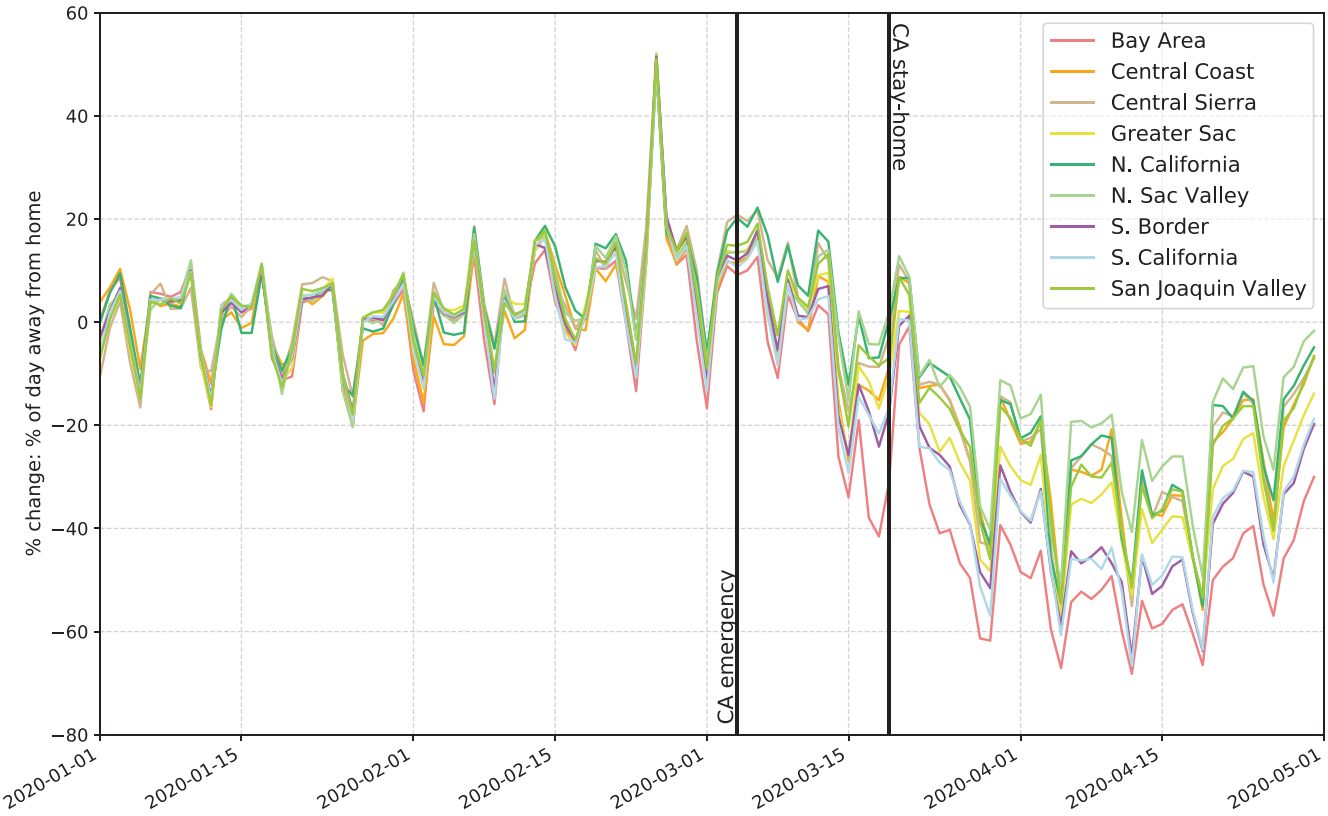
Extended Data Fig. 6 | Choice of, and sensitivity to, weather controls specification. A Sensitivity of main results (Fig. 3) to different functional forms of climate variables (temperature, precipitation, and relative humidity). **B** Mean squared error (MSE) of models fit with different functional forms of weather controls and no other covariates, cross-validated, for both $PM_{2.5}$ and NO_2 models (see Supplementary Information text ('Selection of weather controls') for additional details.). Tables S9-S10 show the MSE statistics for each specification, and the number of observations dropped due to unique weather controls values. We selected interacted vigintile fixed effects as the optimal structure for this analysis based on the combination of lowest MSE for $PM_{2.5}$ and very few dropped observations, even with full interactions of temperature, precipitation, and relative humidity bins.



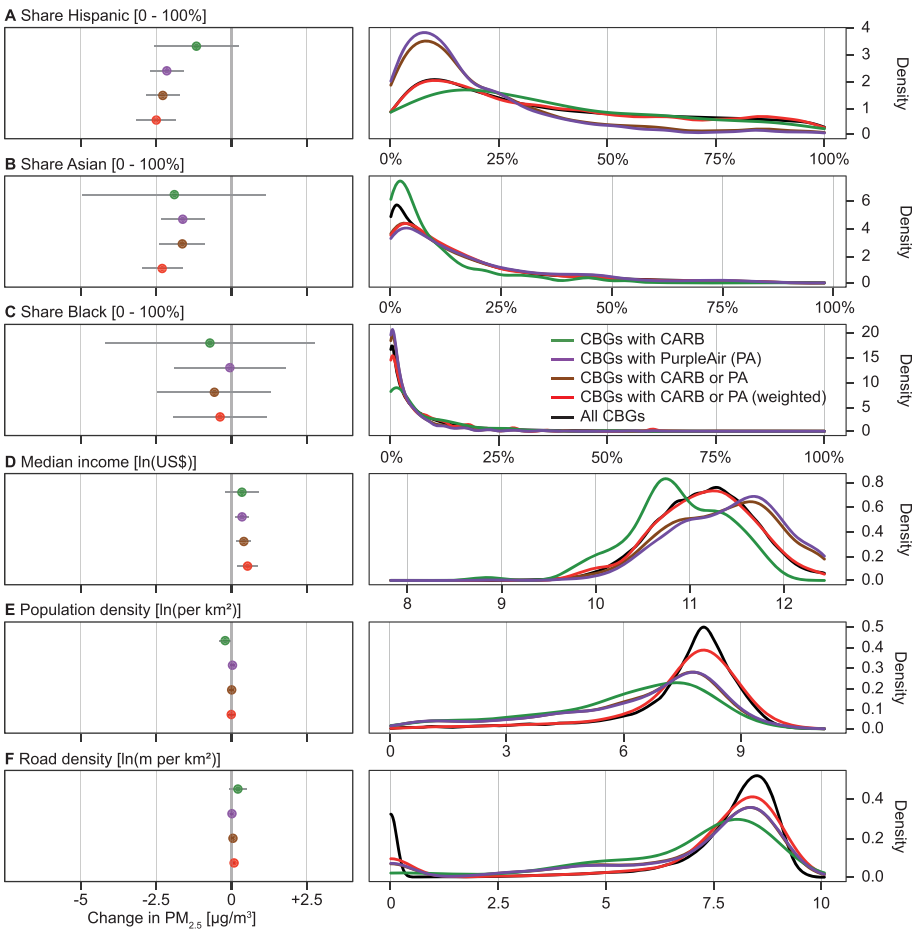
Extended Data Fig. 7 | Daily mean $PM_{2.5}$ concentration summary. Daily measurements from CARB and PurpleAir sensor networks during the study period, in 2019 and 2020, after quality control filtering.



Extended Data Fig. 8 | Map of the number of observations in the TROPOMI NO_2 product. For the time periods pre-shutdown and shutdown for 2019 and 2020 the numbers of observations are shown for each pixel (1 km resolution). Note that the pre-shutdown period is longer than the shutdown period.



Extended Data Fig. 9 | Regional mobility changes in CA. Percent change of mobility is shown for nine regions of CA relative to the mean mobility of each region in January 2020. The urban regions like the Bay Area, Southern CA, and the Southern Border show the largest mobility decrease during the shutdown. The onset of mobility decline occurred essentially simultaneously throughout the state.



Extended Data Fig. 10 | Monitor locations, weighting, and influence on impact estimates. The public (CARB) and private (PurpleAir) $PM_{2.5}$ sensor networks used in this study are not evenly distributed across the state, which affects how different census block groups contribute to estimated impacts. On the left we show post-shutdown concentration changes across various census block group gradients (as in Fig. 3), but estimated using different samples – the public CARB network only (green), the private PurpleAir network only (purple), both together but unweighted (brown), and both together and weighted (red). (These weighted estimates correspond to the estimates presented in Fig. 3.) The panels on the right show the representation of demographic and geographic features due to sensor placement within the different sensor networks. Compared to the distribution of these features by all census block groups in California (black lines), the distribution of census block groups with CARB or PurpleAir monitors can be quite different. The distribution of CARB and PurpleAir combined after weighting (red) matches the all-group state-wide distribution much more closely (see Supplementary Information for details).

nature portfolio

https://doi.org/10.1038/s41893-022-00856-1

Supplementary information

Disparate air pollution reductions during California's COVID-19 economic shutdown

In the format provided by the authors and unedited

Supplementary Information

Disparate air pollution reductions during California's COVID-19 economic shutdown

Richard Bluhm[1,2+], Pascal Polonik[3+], Kyle S. Hemes[4+], Luke C. Sanford[2,5,6+], Susanne A. Benz[6,7+], Morgan C. Levy[3,6+], Katharine Ricke[3,6], Jennifer A. Burney[6*]

[1]School of Economics and Management, Leibniz University Hannover, Germany
[2]Department of Political Science, UC San Diego, USA
[3]Scripps Institution of Oceanography, UC San Diego, USA
[4]Woods Institute for the Environment, Stanford University, USA
[5]School of the Environment, Yale University, USA
[6]School of Global Policy and Strategy, UC San Diego, USA
[7]Centre for Water Resources Studies, Dalhousie University, Canada

[+]These authors contributed equally.
[*]Communicating author: jburney@ucsd.edu

This supplement contains:

- Supplementary Text (7 pages)
- Supplementary Figures S1-S2
- Supplementary Tables S1-S10

Supplementary Text

Data and Methodology Details

California as a study location. California is uniquely well-suited context for this study: it is the fifth largest economy in the world (1, 2); it is one of the most racially and ethnically diverse states in the country (3), and one of only a handful in which non-Hispanic Whites make up less than half the population (4); it is home to four of the top 20 most populous U.S. cities (5); and despite improvements in air quality in the late 1900s and early 2000s (6, 7), several California cities still regularly rank as having the most polluted air in the United States (8). Finally, California citizens live under a historically rich tapestry of environmental regulation – from the Clean Air Act and its amendments at the federal level to local district level rules – that control air pollution from essentially every source in the state. California also has a long history of environmental activism by and on behalf of disadvantaged communities, which have historically experienced higher pollution exposures (9). As such, it is a favorable location for trying to tease apart environmental racism from the legacy of economic policy and other confounds that might also lead to disparate environmental exposures.

However, we note that the use of California as a study region makes interpretation of our results more straightforward than might be the case in other regions, or over larger spatial scales. First, California's mild climate and predictable seasonality makes it easier to compare two years of observations than would be the case in more variable climates. Second, the lack of coal and fuel heating oil use in California means that the regional (anthropogenic) aerosol chemistry is relatively simple – California's $PM_{2.5}$ includes primary carbonaceous aerosols produced by transportation, and secondary nitrates produced by transportation and agriculture (10). There are relatively few other primary sources of particulate matter in California compared to other regions, particularly outside of the state's summer-fall wildfire season, which contributes a large organic carbon burden to the region (11). Our study location and timing also mean that satellite-based NO_2 observations are more highly correlated with $PM_{2.5}$ than they would be in other locations, because the same emissions sources contribute to both in the state (predominantly transportation and agriculture). Studies in more complicated climates, and with a more diverse set of aerosol particulate matter (and precursor) emissions will potentially require more sophisticated statistical techniques to address potential unobserved sources of heterogeneity and to assess whether changes in pollution chemical

composition differ across population subgroups.

Overview of why the COVID-19 related economic slowdown offers new insight into questions of environmental justice. Extended Data Figure 1 compares several quantitative approaches to questions of environmental justice present in the literature. Many environmental justice studies note, as in Extended Data Figure 1A or B, that at any given moment in time (a cross-sectional analysis), ambient pollutant concentrations are higher for communities of color. Here, a best-fit line to cross-sectional observations would lead to an estimate of Δ, or the expected difference in exposure between a 100% Hispanic community and a 100% non-Hispanic community. Accounting for slower-moving confounds in a multi-dimensional analysis, as in B, can change the estimate of Δ. In the case shown, accounting for income can increase the estimate of Δ, if Hispanic households tend to have lower incomes than non-Hispanic households. Many time-varying factors can also confound this relationship. Importantly, expanding to panel (observations across time) analysis, as in Extended Data Figure 1C, allows inclusion of weather variables, and various time cycles known to contribute to changes in pollution, like day-of-week and seasonal effects.

While panel studies allow for inclusion of time-varying covariates, it is still the case that the economy (including both point and mobile sources that emit pollutants like primary PM and other precursors that contribute to secondary PM formation), geography (where humans live, including factors like population density and proximity to roads and other steady-state emissions locations), and climate (annual weather cycle and associated daily and seasonal emissions) typically exist together over a fairly narrow set of conditions. Populations change slowly over time, as does the general structure of the economy. As such, even in panel analyses, it remains difficult to account for enough factors such that residual exposure disparities can be confidently attributed to the broader scale economy.

A large perturbation to the system, as the COVID-19 pandemic has created, moves one piece of the system (the local and non-local in-person economy) far outside the historical experienced conditions. This allows for a much more robust attribution of the change between pre- and post- slowdown conditions to economic factors. The ability to additionally account for 'own' (or local) mobility further allows disaggregation of experienced disparities into those that might be caused by geographic conditions (e.g., communities of color may need to commute more in general, or may be more likely to be essential workers who cannot work from home) and general influence of the broader scale economy. We explain below how this intuition also maps to a statistically well-identified question.

Identification of heterogeneous treatment effects. Consider a simplified version of our specifications in the main text with a single interaction of the treatment status (post-shutdown), d_{it}, with a binary, cross-sectional measure of differential exposure to the treatment, x_i:

$$y_{it} = \tau d_{it} + \gamma(d_{it} \times x_i) + z_{it} + \mu_i + \lambda_t + u_{it}. \tag{1}$$

where z_{it} captures the effects of weather and other observed unit by time variation. All else is defined as before, but note that y_{it} is now in levels to further simplify the exposition. Using annual differences does not fundamentally alter these results but changes all difference-in-differences (DID) interpretations to a triple DID which allows for more complex forms of unobserved heterogeneity.

This set-up allows us to make two points:

1) With staggered treatment ($d_{it} \neq d_t$ for all i) and no heterogeneity in the treatment effect ($\gamma = 0$), $\hat{\tau}$ is a standard DID estimate. Differencing over time, rearranging and taking expectation delivers

$$\hat{\tau} = \mathbb{E}[\Delta y_{it} | \Delta d_{it} = 1] - \mathbb{E}[\Delta y_{it} | \Delta d_{it} = 0] = (\tau + \Delta z_{it} + \Delta\lambda_t) - (\Delta z_{it} + \Delta\lambda_t) = \tau \tag{2}$$

which can be written as $\Delta\bar{y}_{treat} - \Delta\bar{y}_{control}$. However, the overall treatment effect τ is not statistically identified in our setting, where all observations after a particular calendar date are treated, so that $d_{it} = d_t$ for all i, and we do not observe any unit with $\Delta d_{it} = 0$. Hence, d_t and λ_t are perfectly collinear. Omitting the time effects, λ_t would lead us to be able to estimate τ but would also lead us to mistakenly attribute state-wide shocks, $\Delta\lambda_t$, to the treatment effect.

2) Without staggered treatment but heterogeneity in the treatment effect ($\gamma \neq 0$), we can only identify the effect of the heterogeneous exposure, γ, relative to some baseline exposure captured by τ. The treatment effect for each observation now becomes $\tau + \gamma x_i$ but τ is still collinear with λ_t. With binary x_i, the heterogeneous treatment effect is

captured by

$$\begin{aligned}\hat{\gamma} &= \mathbb{E}[\Delta y_{it}|\Delta d_{it}=1, x_i=1] - \mathbb{E}[\Delta y_{it}|\Delta d_{it}=1, x_i=0] \\ &\quad - \{\mathbb{E}[\Delta y_{it}|\Delta d_{it}=0, x_i=1] - \mathbb{E}[\Delta y_{it}|\Delta d_{it}=0, x_i=0]\} \\ &= \mathbb{E}[\Delta y_{it}|\Delta d_{it}=1, x_i=1] - \mathbb{E}[\Delta y_{it}|\Delta d_{it}=1, x_i=0] \\ &= (\tau+\gamma+z_{it}+\Delta\lambda_t) - (\tau+z_{it}+\Delta\lambda_t) = \gamma\end{aligned}$$

which is another DID estimate that compares the effect of the treatment in groups with $x_i = 1$ to those with $x_i = 0$. With continuous x_i this becomes a generalized DID estimate, where we, for example, compare the effect of the shutdown in block groups with a positive share of the Hispanic population to those without any Hispanic residents.

Sensor placement and weighting. Unbiased estimates of experienced pollution changes require that the sample of observations be random and representative. Yet it is well-understood that ground-based monitors might be placed in a non-representative sub-sample of census block groups, as government-funded CARB stations are relatively sparse, and PurpleAir monitors are privately purchased and installed. We find that CARB monitor placement (intentionally) oversamples California's disadvantaged communities – these public monitors are more likely to be placed in poorer, more rural, and more racially and ethnically diverse neighborhoods (Extended Data Figure 10A-F). The PurpleAir network is unsurprisingly slanted towards wealthier locations and under-represents the Hispanic population of California (Extended Data Figure 10A-F). While both do not reflect the true distribution of population characteristics, the sheer size of the Purple Air networks implies that it spans a large variety of communities.

Choice-based sampling implies that monitor placement is correlated with the error term of our regression equation. Estimation which ignores endogenous sampling is generally biased, but consistent estimates can be obtained by weighting the regression function with the inverse probability of selection (12). We use iterative proportional fitting—a standard post-stratification procedure—to match the marginal distributions of the endogenous sample of monitors to known census population margins. We determine the marginal distributions of the population by computing the cell frequencies for each of the more detailed ACS variables used in the analysis (the Hispanic, Asian, and Black population shares, as well as income, road density and population density) per vigintile (20 bins) of the census data. A process called 'raking' then finds post-stratification weights which adjust the endogenous sample such that it resembles the set of target distributions (Extended Data Figure 10). The process is able to fit individual distributions

in our data very well, but involves some trade-offs in terms of how well it matches any particular distribution when more than one target variable is used. Calculating the inverse probability of selection directly on a large contingency table would require a much coarser portioning of the data, resulting in a substantially worse fit. All fixed effects regressions in the main text which use ground-based monitors are estimated using weighted least squares if not otherwise noted.

We re-estimate impacts using several subsets of census block groups, including only CARB monitors, only PurpleAir monitors, and both networks combined with and without weighting (Extended Data Figure 10). While the racial disparities estimated in the weighted sample are consistent with most unweighted samples (in the sense that the uncertainties overlap), there are important gains from using both networks together and weighting the observations. The CARB-only estimates fail to detect significant differences in pollution for most variables, including income. The bias towards more rural areas in the state's CARB network manifests itself in large estimates for the effect of road density on air pollution disparities. The unweighted distribution of the PurpleAir network is usually closer to the distribution of the underlying population characteristic in the ACS data. The sampled locations closely approximate the population distribution only once both networks are combined and weighted. Using the weights increases the (absolute) size of the pollution disparity estimated for Asian populations by about 41% and the effect of income by 29%.

We do not report results where we separately derive weights for the census block groups cover by either the CARB or PurpleAir sensors. The selective placement of CARB sensors together with their low number of observations in particular makes it impossible to derive weights which "undo" the over-representation of disadvantaged communities (additional results are available on request). We need both systems to cover the joint distribution of race/ethnicity, income, population density and road density in California. We can, however, use the full distribution of weights to investigate what these weights tell us about the potential for each set of monitors to provide information on disparities. Figure 2C shows the distribution of weights across CARB and PurpleAir sensor types and shows that more CARB sensors are highly upweighted by our weighting process than PurpleAir sensors. This makes sense–given the endogenous placement of sensors discussed above and in the main text–we should expect some CARB sensors to be deliberately placed where there are few other monitors or in locations where residents would be unlikely to purchase a PurpleAir sensor.

We note that unlike the surface $PM_{2.5}$ networks, NO_2 satellite data cover the entire state and are thus perfectly representative. However, a remaining potential sampling issue with satellite data are biases related to missing data (for example, due to cloud screening in rainy seasons). Extended Data Figure 8 shows the distribution of observations in the satellite NO_2 record. Some pixels (1km) have have only a few observations, particularly in the pre-shutdown winter period, and areas that tend to have cloudcover later in the spring (e.g., the Sierra Nevada range) also have more missing observations. However because data are primarily missing for rural areas, this only translates into 5 census block group with available demographics but incomplete weekly NO_2 observations.

Selection of weather controls. We have included variables to control for temperature, relative humidity, and precipitation in the same day ($PM_{2.5}$) or week (NO_2) as the dependent variable measurement. Most importantly, these controls help reduce the variability caused by weather differences between 2019 and 2020. For example, if areas with larger Asian populations also received more rainfall in 2020 after the shutdown, we might mis-attribute the subsequent reduction in $PM_{2.5}$ concentrations to the shutdown rather than to the weather.

The form of the function which maps these three variables onto concentrations is unknown to us, so we searched for a specification which had good out-of-sample performance, didn't use too many degrees of freedom, and where interacted fixed effects specifications contained few bins with only one observation (effectively dummying out that observation). This exercise sought to find a well-fitting but not overfit specification that left enough observations for the main models to work well, and that did not accidentally remove observations from geographic areas with extremes in one or more of the variables.

Tables S9 and S10 as well as Extended Data Figure 6 show the results of these specification searches. The categories of functional forms that we tried included interacted fixed-width fixed effects (one, two, five, ten, twenty) of native units, fixed effects created to split the data into evenly sized groups (decile, vigintile), a cubic spline, polynomial fits (first, second, and third degree), uninteracted fixed effects, and a specification with no adjustment. In the tables, we show each of the first four classes of models using both 2019 and 2020 weather variables or only 2020 weather variables, meaning we included one or both of $f^{2020}(T, RH, P)_{it} + f^{2019}(T, RH, P)_{it}$. To test the performance of each specification we ran a regression of weather variables on pollutant using a randomly chosen 70% of CBGs and tested the performance on the remaining 30%. Tables S9 and S10 show the average out of sample mean squared error (MSE) for each specification across 100 random

splits, the standard deviation of those estimates (Std. Dev), the degrees of freedom used by each specification (DoF), and the number of observations which are dummied out because they singularly correspond to a fixed effect (Lost). Both tables are ranked from lowest to highest MSE.

Extended Data Figure 6B shows the distribution of MSE estimates across 100 70%/30% cross-validations for each specification for $PM_{2.5}$ and NO_2. Panel A shows the coefficients from our main model using a subset of the weather specifications. Our preferred specification is vigintile fixed effects, interacted, for both 2019 and 2020. This specification minimized MSE in the $PM_{2.5}$ regressions and was second in NO_2 regressions, while drastically reducing the degrees of freedom used and the observations dummied out.

Race and ethnic group aggregation. There is important variation within racial and ethnic groups not represented by the group aggregations evaluated in this study. Racial and ethnic groups were not disaggregated into subgroups because of data availability and methodological suitability. Asian subgroup identification, for example, is not available in the ACS at the CBG level, only at the tract level. Even when summarizing Asian subgroups at the tract level, disaggregated population counts are small. Even the largest Asian subgroups (Filipino and Chinese) would have fewer samples than our current smallest primary group (Black). Therefore, Asian subgroup analyses would face sampling issues from which we would be unable to draw clear conclusions using the methods in this study. We already faced a similar issue with respect to sample size for Black populations (see main text Results). Analyses of subgroup dynamics would be valuable, but our study does not employ a design appropriate for that investigation.

Additional discussion of mobility results

Consistent with other research (13–16), we find large differences in mobility across different income and racial groups. Census block groups with high Hispanic and Black populations had smaller mobility reductions during the shutdown than predominantly non-Hispanic White neighborhoods (0.8 pp for every 10 pp increase in the population share). However, these differences can be completely accounted for by allowing for heterogeneous effects in income. This suggests that mobility during the pandemic is mainly a function of the economic ability to stay home and the probability of belonging to the essential workforce, rather than other characteristics associated with different neighborhoods. This does not hold for block groups with a greater share of the Asian population. Here we

estimate a 0.20 pp decrease in mobility for a 10 pp increase in the census block group Asian population share. The effect falls to -0.14 pp but remains highly significant even after allowing for heterogeneous responses to other block group characteristics.

References

[1] U.S. Bureau of Economic Analysis (BEA). Gross Domestic Product by State: 4th Quarter and Annual 2019.

[2] International Monetary Fund. Report for Selected Countries and Subjects, October 2019.

[3] U.S. News and World Reports. Majority of U.S. Cities are Becoming More Diverse, New Analysis Shows, January 2020.

[4] US Census Bureau. ACS Provides New State and Local Income, Poverty and Health Insurance Statistics, September 2019.

[5] US Census Bureau. City and Town Population Totals: 2010-2019, May 2020.

[6] David D Parrish, Jin Xu, Bart Croes, and Min Shao. Air quality improvement in los angeles—perspectives for developing cities. *Frontiers of Environmental Science & Engineering*, 10(5):11, 2016.

[7] Erika Garcia, Robert Urman, Kiros Berhane, Rob McConnell, and Frank Gilliland. Effects of policy-driven hypothetical air pollutant interventions on childhood asthma incidence in southern California. *Proceedings of the National Academy of Sciences*, 116(32):15883–15888, 2019.

[8] American Lung Association. State of the Air 2020. Technical report, American Lung Association, Chicago, Illinois, 2020.

[9] Manuel Pastor Jr, Rachel Morello-Frosch, and James L Sadd. The air is always cleaner on the other side: Race, space, and ambient air toxics exposures in California. *Journal of Urban Affairs*, 27(2):127–148, 2005.

[10] S. Hasheminassab, N. Daher, A. Saffari, D. Wang, B. D. Ostro, and C. Sioutas. Spatial and temporal variability of sources of ambient fine particulate matter ($PM_{2.5}$) in California. *Atmospheric Chemistry and Physics*, 14(22):12085–12097, 2014.

[11] Dan Jaffe, William Hafner, Duli Chand, Anthony Westerling, and Dominick Spracklen. Interannual variations in $PM_{2.5}$ due to wildfires in the Western United States. *Environmental Science & Technology*, 42(8):2812–2818, 2008.

[12] Jeffrey M. Wooldridge. Asymptotic properties of weighted M-Estimators for variable probability samples. *Econometrica*, 67(6):1385–1406, 1999.

[13] Giovanni Bonaccorsi, Francesco Pierri, Matteo Cinelli, Andrea Flori, Alessandro Galeazzi, Francesco Porcelli, Ana Lucia Schmidt, Carlo Michele Valensise, Antonio Scala, Walter Quattrociocchi, et al. Economic and social consequences of human mobility restrictions under Covid-19. *Proceedings of the National Academy of Sciences*, 117(27):15530–15535, 2020.

[14] Michael S Warren and Samuel W Skillman. Mobility changes in response to Covid-19. *arXiv preprint arXiv:2003.14228*, 2020.

[15] Caroline O Buckee, Satchit Balsari, Jennifer Chan, Mercè Crosas, Francesca Dominici, Urs Gasser, Yonatan H Grad, Bryan Grenfell, M Elizabeth Halloran, Moritz UG Kraemer, et al. Aggregated mobility data could help fight Covid-19. *Science*, 368(6487):145, 2020.

[16] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of Covid-19 explain inequities and inform reopening. *Nature*, pages 1–6, 2020.

Supplementary Figures

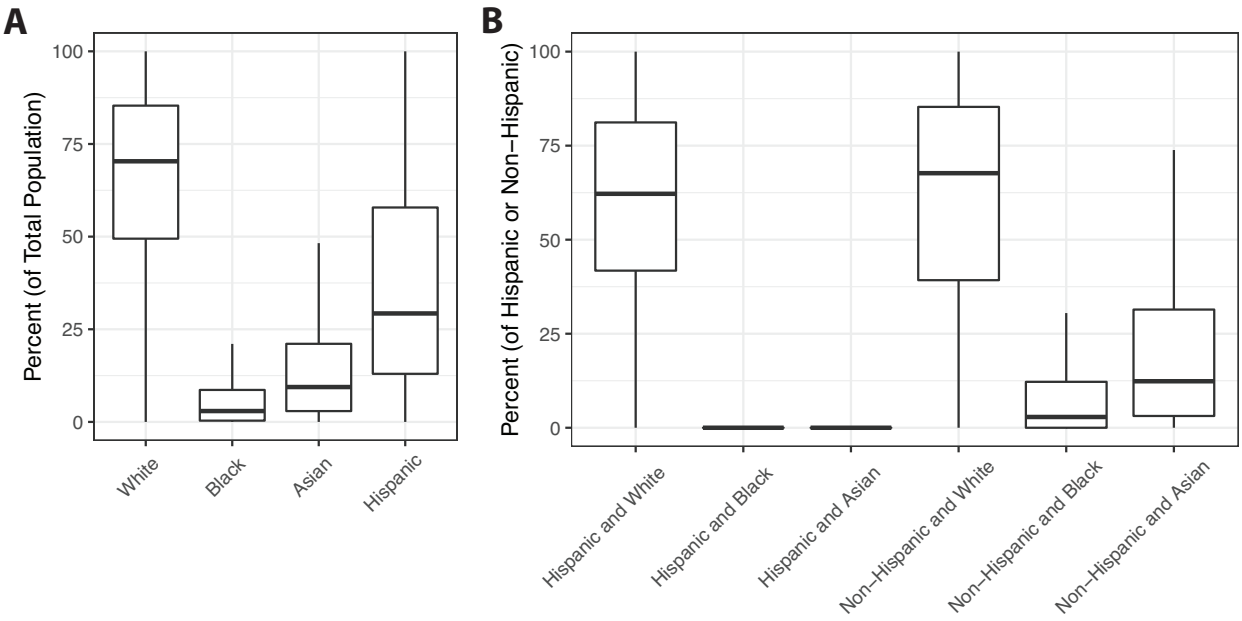


Figure S1: **Summary of Hispanic ethnicity and race for California census block groups. A** Distributions of the percentages of the total population categorized as White, Black, Asian, and Hispanic, and **B** the percentages of either the Hispanic or non-Hispanic portion of the total population categorized as White, Black, and Asian at the census block group (CBG) level according to the U.S. Census Bureau 2018 5-year American Community Survey (ACS). The total number of CBGs was 23,212. The boxplots display the median (50th percentile), two hinges (25th and 75th percentiles), and two whiskers (largest value $\leq$ 1.5*IQR; IQR is the distance between hinges); statistical outliers were included in the distribution calculation, but excluded from the visualization.

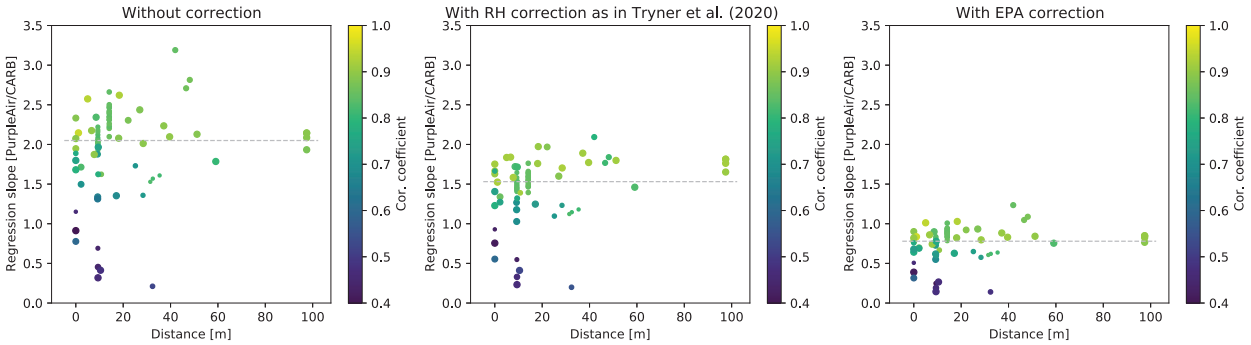


Figure S2: **Regression slopes of nearly co-located (<100m) PurpleAir and CARB $PM_{2.5}$ measurements (2019 and 2020 Jan-Apr outdoor in CA) as a function of distance between the PurpleAir and CARB sensors.** Color indicates the correlation coefficient. Marker size indicates the number of points used for the regression. The dashed horizontal line indicates the median. Point size indicates number of points available for the sensor pairs. The three panels show the same comparison with no correction (left), the RH correction proposed in Tryner et al. (2020), and the EPA correction used in the study (right). The RH correction may not be sufficient in our case because we use daily mean data. The EPA correction we implement is conservative. Several PurpleAir sensors are paired with the same CARB sensors (in one case as many as 32). 29 unique CARB sensors are represented here, meaning they have a PurpleAir sensor within 100m. Three sensor pairs were excluded from the figure because the correlation was very low (< 0.2).

Supplementary Tables

Racial/Ethnic category	Percentile	Pre-shutdown	Post-shutdown	% change
Hispanic	10	46.7	19.0	-59.4
Hispanic	90	53.6	39.1	-27.1
Asian	10	54.1	40.0	-26.1
Asian	90	46.3	18.6	-59.8
Black	10	48.7	29.2	-40.0
Black	90	54.0	38.4	-28.9

Table S1: Average pre-/post-shutdown percent of time spent away from home for the 10th and 90th percentile CBG of population percentage of each group. For example, the first row represents the CBG at the tenth percentile of Hispanic population proportion, which went from 46.7% to 19.0% of their time spent away from home, representing a 59.4 percent decrease in mobility. This table summarizes Figure 2B in the main text.

	Dependent variable: Difference in $PM_{2.5}$ [μg m^{-3}]										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Mobility	−0.421					−0.321				0.260	−0.119
	(0.159)**					(0.162)*				(0.220)	(0.160)
Income		0.992					0.587			0.903	0.526
		(0.171)***					(0.158)***			(0.251)***	(0.163)***
Road Density			0.007					0.017		0.102	−0.002
			(0.017)					(0.017)		(0.040)**	(0.018)
Population Density				−0.018					0.120	−0.104	0.087
				(0.059)					(0.055)**	(0.119)	(0.053)
% non-White and/or Hispanic					−2.566	−2.549	−2.102	−2.574	−2.771	−2.956	−2.291
					(0.216)***	(0.217)***	(0.268)***	(0.214)***	(0.230)***	(0.814)***	(0.279)***
CBG FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Day FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$f^{2019}(T, RH, P)$	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓
$f^{2020}(T, RH, P)$	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓
Observations	73,255	73,255	73,255	73,255	73,255	73,255	73,255	73,255	73,255	73,255	73,255
Note:											*p<0.1; **p<0.05; ***p<0.01

Table S2: Results of our $PM_{2.5}$ regressions compare the post-shutdown difference from 2020 to 2019 to the pre-shutdown difference. The first four columns correspond to Figure 3A, all other columns to Figure 3C. Mobility is represented as percentage of time spend away from home and coefficients are the estimated difference between 0 and 100%. The coefficients for income, road density and population density correspond to one logarithmic unit, while the coefficients for our demographic variables correspond to changes between 0 and 100% population share. Regressions focusing on the demographic variables (columns 5 to 11) are run for seven models: no control variables, controlling for mobility, controlling for income, controlling for road density, controlling for population density, controlling for everything but excluding weather effects, and controlling for all and weather effects. Standard errors clustered at the county level are shown in parentheses below point estimates.

	Dependent variable: Difference in $PM_{2.5}$ [μg m^{-3}]						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mobility		−0.277				0.217	−0.128
		(0.154)*				(0.203)	(0.160)
Income			0.582			0.987	0.535
			(0.157)***			(0.223)***	(0.164)***
Road Density				0.009		0.096	−0.006
				(0.018)		(0.040)**	(0.019)
Population Density					0.109	−0.101	0.088
					(0.053)**	(0.114)	(0.052)*
% Hispanic	−2.858	−2.835	−2.333	−2.858	−3.017	−3.190	−2.493
	(0.257)***	(0.257)***	(0.314)***	(0.257)***	(0.271)***	(0.711)***	(0.331)***
% Asian	−1.885	−1.941	−2.028	−1.903	−2.218	−3.435	−2.298
	(0.498)***	(0.483)***	(0.390)***	(0.505)***	(0.416)***	(1.075)***	(0.340)***
% Black	−0.623	−0.597	−0.168	−0.635	−0.862	1.062	−0.377
	(0.819)	(0.816)	(0.799)	(0.811)	(0.801)	(1.196)	(0.793)
CBG FEs	✓	✓	✓	✓	✓	✓	✓
Day FEs	✓	✓	✓	✓	✓	✓	✓
$f^{2019}(T, RH, P)$	✓	✓	✓	✓	✓	–	✓
$f^{2020}(T, RH, P)$	✓	✓	✓	✓	✓	–	✓
Observations	73,255	73,255	73,255	73,255	73,255	73,255	73,255
Note:							*p<0.1; **p<0.05; ***p<0.01

Table S3: Results of our $PM_{2.5}$ regressions compare the post-shutdown difference from 2020 to 2019 to the pre-shutdown difference. The columns correspond to Figure 3E. Mobility is represented as percentage of time spend away from home and coefficients are the estimated difference between 0 and 100%. The coefficients for income, road density and population density correspond to one logarithmic unit, while the coefficients for our demographic variables correspond to changes between 0 and 100% population share. The regressions are run for seven models: no control variables, controlling for mobility, controlling for income, controlling for road density, controlling for population density, controlling for everything but excluding weather effects, and controlling for all and weather effects. Standard errors clustered at the county level are shown in parentheses below point estimates.

	Dependent variable: Difference in Mobility (% away from home)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Income	−0.121				−0.130			−0.127	−0.124
	(0.010)***				(0.009)***			(0.011)***	(0.010)***
Road Density		−0.005				−0.005		−0.001	−0.002
		(0.002)**				(0.002)***		(0.002)	(0.002)
Population Density			−0.013				−0.017	−0.010	−0.009
			(0.002)***				(0.003)***	(0.003)***	(0.003)***
% non-White and/or Hispanic				0.055	−0.047	0.058	0.085	−0.045	−0.026
				(0.018)***	(0.017)***	(0.018)***	(0.018)***	(0.026)*	(0.019)
CBG FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Day FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
$f^{2019}(T, RH, P)$	✓	✓	✓	✓	✓	✓	✓	–	✓
$f^{2020}(T, RH, P)$	✓	✓	✓	✓	✓	✓	✓	–	✓
Observations	73,255	73,255	73,255	73,255	73,255	73,255	73,255	73,255	73,255
Note:									*p<0.1; **p<0.05; ***p<0.01

Table S4: Results of our mobility regressions compare the post-shutdown difference from 2020 to 2019 to the pre-shutdown difference. The first three columns correspond to Figure 3B, all other columns to Figure 3D. The coefficients for income, road density and population density correspond to one logarithmic unit, while the coefficients for our demographic variables correspond to changes between 0 and 100% population share. Regressions focusing on the demographic variables (columns 4 to 9) are run for six models: no control variables, controlling for income, controlling for road density, controlling for population density, controlling for everything but excluding weather effects, and controlling for all and weather effects. Standard errors clustered at the county level are shown in parentheses below point estimates.

	Dependent variable: Difference in Mobility (% away from home)					
	(1)	(2)	(3)	(4)	(5)	(6)
Income		−0.109			−0.108	−0.106
		(0.010)***			(0.011)***	(0.011)***
Road Density			−0.004		−0.001	−0.002
			(0.002)**		(0.002)	(0.002)
Population Density				−0.012	−0.007	−0.007
				(0.003)***	(0.003)**	(0.003)**
% Hispanic	0.084	−0.015	0.084	0.101	−0.020	−0.001
	(0.022)***	(0.025)	(0.022)***	(0.021)***	(0.034)	(0.025)
% Asian	−0.200	−0.173	−0.193	−0.164	−0.168	−0.149
	(0.054)***	(0.025)***	(0.053)***	(0.052)***	(0.028)***	(0.027)***
% Black	0.093	0.008	0.099	0.120	0.009	0.029
	(0.047)*	(0.049)	(0.047)**	(0.047)**	(0.055)	(0.051)
CBG FEs	✓	✓	✓	✓	✓	✓
Day FEs	✓	✓	✓	✓	✓	✓
$f^{2019}(T, RH, P)$	✓	✓	✓	✓	–	✓
$f^{2020}(T, RH, P)$	✓	✓	✓	✓	–	✓
Observations	73,255	73,255	73,255	73,255	73,255	73,255

Note: *p<0.1; **p<0.05; ***p<0.01

Table S5: Results of our mobility regressions compare the post-shutdown difference from 2020 to 2019 to the pre-shutdown difference. The columns correspond to Figure 3F. The coefficients for income, road density and population density correspond to one logarithmic unit, while the coefficients for our demographic variables correspond to changes between 0 and 100% population share. The regressions are run for six models: no control variables, controlling for income, controlling for road density, controlling for population density, controlling for everything but excluding weather effects, and controlling for all and weather effects. Standard errors clustered at the county level are shown in parentheses below point estimates.

	Dependent variable: NO_2 [μmol/m^2]										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Mobility	1.318					2.234				9.232	2.968
	(1.957)					(1.994)				(3.833)**	(1.128)**
Income		3.244					2.123			−4.934	1.901
		(2.559)					(3.124)			(3.884)	(2.840)
Road Density			0.245					0.240		2.737	0.373
			(0.206)					(0.197)		(0.634)***	(0.081)***
Population Density				−1.753					−1.503	−6.643	−1.629
				(0.409)***					(0.400)***	(1.988)***	(0.360)***
% non-White and/or Hispanic					−7.378	−7.588	−5.489	−7.354	−4.306	−28.596	−2.599
					(2.469)***	(2.532)***	(3.986)	(2.448)***	(2.504)*	(6.682)***	(3.887)
CBG FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Day FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$f^{2019}(T, RH, P)$	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓
$f^{2020}(T, RH, P)$	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓
Observations	370,378	370,378	370,378	370,378	370,378	370,378	370,378	370,378	370,378	370,379	370,378

Note: *p<0.1; **p<0.05; ***p<0.01

Table S6: Results of our weekly NO_2 regressions compare the post-shutdown difference from 2020 to 2019 to the pre-shutdown difference. The first four columns correspond to Extended Data Figure 4A, all other columns to Extended Data Figure 4C. Mobility is represented as percentage of time spend away from home and coefficients are the estimated difference between 0 and 100%. The coefficients for income, road density and population density correspond to one logarithmic unit, while the coefficients for our demographic variables correspond to changes between 0 and 100% population share. Regressions focusing on the demographic variables (columns 5 to 11) are run for seven models: no control variables, controlling for mobility, controlling for income, controlling for road density, controlling for population density, controlling for everything but excluding weather effects, and controlling for all and weather effects. Standard errors clustered at the county level are shown in parentheses below point estimates.

	Dependent variable: NO_2 [μmol/m^2]						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mobility		0.923				8.722	2.298
		(2.055)				(4.101)**	(1.232)*
Income			4.020			−3.143	3.669
			(3.336)			(5.217)	(3.042)
Road Density				0.328		2.788	0.417
				(0.195)*		(0.671)***	(0.086)***
Population Density					−1.468	−6.478	−1.588
					(0.413)***	(1.951)***	(0.345)***
% Hispanic	−6.433	−6.535	−2.579	−6.305	−3.489	−27.299	0.179
	(2.725)**	(2.690)**	(3.194)	(2.636)**	(2.554)	(9.195)***	(3.194)
% Asian	−20.406	−20.363	−20.024	−20.912	−17.146	−38.339	−17.067
	(3.967)***	(3.980)***	(4.032)***	(4.104)***	(3.872)***	(12.817)***	(4.005)***
% Black	3.973	3.865	8.142	3.855	7.254	−11.928	10.910
	(7.627)	(7.788)	(10.244)	(7.599)	(7.776)	(6.480)*	(9.914)
CBG FEs	✓	✓	✓	✓	✓	✓	✓
Day FEs	✓	✓	✓	✓	✓	✓	✓
$f^{2019}(T, RH, P)$	✓	✓	✓	✓	✓	–	✓
$f^{2020}(T, RH, P)$	✓	✓	✓	✓	✓	–	✓
Observations	370,378	370,378	370,378	370,378	370,378	370,379	370,378
Note:							*p<0.1; **p<0.05; ***p<0.01

Table S7: Results of our weekly NO_2 regressions compare the post-shutdown difference from 2020 to 2019 to the pre-shutdown difference. The columns correspond to Extended Data Figure 4E. Mobility is represented as percentage of time spend away from home and coefficients are the estimated difference between 0 and 100%. The coefficients for income, road density and population density correspond to one logarithmic unit, while the coefficients for our demographic variables correspond to changes between 0 and 100% population share. The regressions are run for seven models: no control variables, controlling for mobility, controlling for income, controlling for road density, controlling for population density, controlling for everything but excluding weather effects, and controlling for all and weather effects. Standard errors clustered at the county level are shown in parentheses below point estimates.

	Dependent variable:					
	Difference in $PM_{2.5}$ [μg m^{-3}]			$PM_{2.5}$ in 2020		
	(1)	(2)	(3)	(4)	(5)	(6)
Mobility		−0.232	−0.128		−0.228	−0.395*
		(0.140)	(0.160)		(0.176)	(0.215)
Income		0.412***	0.535***		1.071***	1.348***
		(0.125)	(0.164)		(0.190)	(0.275)
Road density		0.008	−0.006		0.013	−0.006
		(0.022)	(0.019)		(0.031)	(0.025)
Population density		0.051	0.088*		−0.229***	−0.108*
		(0.039)	(0.052)		(0.048)	(0.058)
% Hispanic	−2.629***	−2.278***	−2.493***	−3.242***	−1.758***	−2.008***
	(0.268)	(0.284)	(0.331)	(0.478)	(0.497)	(0.567)
% Asian	−0.984*	−1.629***	−2.298***	−0.536	−0.544	−1.545**
	(0.564)	(0.386)	(0.340)	(0.869)	(0.556)	(0.599)
% Black	−0.642	−0.565	−0.377	−2.251	−0.568	−1.521
	(0.965)	(0.958)	(0.793)	(1.404)	(1.246)	(1.429)
Weighted	–	–	✓	–	–	✓
CBG FEs	✓	✓	✓	✓	✓	✓
Day FEs	✓	✓	✓	✓	✓	✓
$f^{2019}(T, RH, P)$	✓	✓	✓	–	–	–
$f^{2020}(T, RH, P)$	✓	✓	✓	✓	✓	✓
Observations	73,255	73,255	73,255	73,255	73,255	73,255
Note:						*p<0.1; **p<0.05; ***p<0.01

Table S8: $PM_{2.5}$ 2020-2019 difference vs 2020 values for all of California. Results of three $PM_{2.5}$ regressions: without controls or weights, with controls and without weights, and with both controls and weights. The first three columns correspond to the estimates in the paper, the second three columns estimate the same quantities using only 2020 data. The dependent variable is the level of $PM_{2.5}$ rather than the inter-annual difference, and mobility corresponds to the level of mobility rather than the difference. We only include 2020 weather variables in columns 3-6 rather than the both years. Standard errors clustered at the county level are shown in parentheses below point estimates.

The estimates have slight differences in magnitude but are the same sign and significance level in both specifications. This has two implications. First, the weather controls do a good enough job that differencing out the previous year does not make an enormous difference, though it does add precision to the estimates. Second, it is possible to consistently estimate these effects without access to previous years data.

Rank	Model	MSE	Std. Dev.	DoF	Lost
1	By vigintile	22.3	1.3	4041	409
2	By decile	22.6	1.4	735	16
3	By 2-unit increment	22.8	1.3	5808	2441
4	By 5-unit increment	23.3	1.4	1078	260
5	Cubic spline	23.7	1.5	22	0
6	Uninteracted	23.8	1.5	480	40
7	By 1-unit increment	23.8	1.1	15597	8969
8	By 10-unit increment	23.9	1.5	269	34
9	By vigintile, 2020	25.2	1.5	1992	152
10	By 2-unit increment, 2020	25.5	1.5	2086	908
11	By 20-unit increment	25.5	1.6	61	6
12	By 1-unit increment, 2020	25.6	1.5	5800	3674
13	By decile, 2020	25.6	1.6	360	10
14	By 5-unit increment, 2020	25.9	1.6	391	91
15	Cubic spline, 2020	26.0	1.6	43	0
16	By 10-unit increment, 2020	26.2	1.6	97	12
17	1st degree polynomial	26.6	1.6	7	0
18	By 20-unit increment, 2020	26.7	1.7	25	3
19	2nd degree polynomial, 2020	27.0	1.5	10	0
20	3rd degree polynomial, 2020	27.1	1.4	20	0
21	1st degree polynomial, 2020	27.9	1.7	4	0
22	base	28.9	1.8	1	0
23	2nd degree polynomial	31.9	2.0	28	0
24	3rd degree polynomial	34.1	2.4	71	0

Table S9: Results of cross-validated $PM_{2.5}$ selection of weather controls according to MSE criterion as shown on the left side of Extended Data Figure 6B. Each row corresponds to one of the specifications we considered for controlling for temperature, humidity, and precipitation. Rows are ordered from smallest to largest out of sample mean squared error (MSE). We also include the degrees of freedom used by the specification as well as the number of observations which cannot be used to estimate the parameters of interest because they uniquely correspond to a fixed effect. Main regressions use the 'By vigintile' specification

Rank	Model	MSE (1e-11)	Std. Dev. (1e-11)	DoF	Lost
1	By 1-unit increment	143.00	3.60	14453	3995
2	By vigintile	178.30	5.10	7467	590
3	By 2-unit increment	209.60	5.70	3844	815
4	By decile	238.60	6.40	1355	25
5	By vigintile, 2020	285.60	7.70	3654	297
6	By 1-unit increment, 2020	289.80	13.00	5419	1363
7	Uninteracted	304.50	8.10	317	8
8	By 5-unit increment	307.00	8.70	523	45
9	By decile, 2020	334.00	9.20	670	13
10	Cubic spline, 2020	342.60	12.20	43	0
11	By 2-unit increment, 2020	371.60	10.60	1318	196
12	By 10-unit increment	372.30	10.50	111	5
13	By 5-unit increment, 2020	401.20	11.70	173	12
14	By 10-unit increment, 2020	420.60	12.00	38	2
15	By 20-unit increment,	425.90	12.00	29	0
16	Cubic spline	436.00	19.00	22	0
17	By 20-unit increment, 2020	438.60	12.50	12	0
18	1st degree polynomial	438.60	12.30	7	0
19	2nd degree polynomial, 2020	455.30	13.40	10	0
20	1st degree polynomial, 2020	465.00	13.70	4	0
21	Base	466.20	13.90	1	0
22	3rd degree polynomial, 2020	478.60	17.60	19	0
23	3rd degree polynomial	738.10	79.20	64	0
24	2nd degree polynomial	818.20	99.00	28	0

Table S10: Results of cross-validated NO_2 selection of weather controls according to MSE criterion as shown on the right side of Extended Data Figure 6B. Each row corresponds to one of the specifications we considered for controlling for temperature, humidity, and precipitation. Rows are ordered from smallest to largest out of sample mean squared error (MSE). We also include the degrees of freedom used by the specification as well as the number of observations which cannot be used to estimate the parameters of interest because they uniquely correspond to a fixed effect. Main regressions use the 'By vigintile' specification because it uses many fewer degrees of freedom and drops fewer observations than the 'By 1' specification, and is consistent with what is used in the $PM_{2.5}$ specification.

nature portfolio

Corresponding author(s): Jennifer Burney

Last updated by author(s): Jan 3, 2022

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	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	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)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	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 were downloaded from public repositories using code written by the authors in R and the Google Colab Python interface with Google Earth Engine.
Data analysis	Software used for analysis included RStudio and R statistical software (version 4.0.x), and Google Colab Notebooks. These codes are available in our group GitHub repository (https://github.com/jaburney/CA-COVIDEJ-2022). We used ArcGIS to generate several of the maps, and Adobe Illustrator to assemble final figures.

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

The pollution, weather, and demographic data used in this project are all publicly available. PurpleAir data are accessible via public API (https://api.purpleair.com/), CARB data can be found at (https://www.arb.ca.gov/aqmis2/aqmis2.php). Gridded NO2 data (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_NRTI_L3_NO2) and GridMet weather data (https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_GRIDMET) are available via Google Earth Engine. Mobility data from Safegraph are available to academic researchers (https://www.safegraph.com/academics) but may not be

reposted in raw format. All publicly-available raw data for this project can be found in our project dataverse (https://doi.org/10.7910/DVN/ZXVB7A), and processed and derived data can be found in our GitHub repository (https://github.com/jaburney/CA-COVIDEJ-2022). Finally, we have provided an online tool to facilitate exploration of the data at the Census Block Group scale (https://sabenz.users.earthengine.app/view/covid-ej).

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

Ecological, evolutionary & environmental sciences study design

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

Study description	This is a retrospective, empirical study that relates ground-based PM2.5 and satellite-based NO2 pollution measures to Census Block Group demographics and uses the COVID-19 shutdown in spring of 2020 to statistically identify disparate pollution reductions across California's largest racial and ethnic minority groups. The panel analysis controls for weather, population density, road density, mobility (defined as the % of the day spent away from home), income, location-specific-time-invariant effects, and temporal effects (day of week, day of year).
Research sample	The sample is present-day California Census Block Groups (CBGs) as defined and measured in the 2018 American Community Survey (ACS). There are 23,212 Census Block Groups in California at this time.
Sampling strategy	We retained all CBGs for which a complete record of observations over time time period existed. For satellite-based NO2 measurements, this was near-full coverage (22,503 CBGs). For ground-based PM2.5 measurements, 733 unique CBGs were represented in the data.
Data collection	PM2.5 Data: Surface station measurements of particulate matter with diameter smaller than 2.5 microns (PM2.5) were downloaded from publicly available databases from PurpleAir and the California Air Resources Board (CARB). We downloaded all outdoor PurpleAir data available (1891 individual stations) for Jan-Apr 2019 and 2020. We retrieved (May 1, 2020) all hourly CARB PM2.5 data in California available for Jan-Apr 2019 and 2020 using CARB's Air Quality and Meteorological Information System (AQMIS), and calculated the daily mean (150 individual stations). NO2 Data: We used the Copernicus Sentinel-5 Precursor TROPOspheric Monitoring Instrument (TROPOMI, version 1.03.02) Offline tropospheric NO2 column number density for mean NO2 concentrations of the developed areas of each census block group. TROPOMI has a resolution of 0.01 arc degrees. Data were collected for Jan-Apr 2019 and 2020 and only for developed areas based on the U.S. Geological Survey (USGS) National Land Cover Database (NLCD) 2016. For this study, all data was prepared using the Google Earth Engine Python API and formatted as weekly means for each census block group. Weekly means were chosen to counteract the high frequency of missing data, particularly in northern California. Climate Data: For temperature, precipitation, and relative humidity we relied on the Gridded Surface Meteorological dataset (GridMet). GridMet provides daily information at 4-km resolution across the continental USA. For this study, data were aggregated in Google Earth Engine in its original daily frequency for each PM2.5 measurement station, and as a weekly mean for the NO2 Data for each census block group. The weekly mean data was only collected for developed areas based on the U.S. Geological Survey (USGS) National Land Cover Database (NLCD) 2016. Mobility Data: We use SafeGraph's Social Distancing Metrics, which were made available for research as part of the company's COVID-19 response, and have been validated elsewhere. SafeGraph collects and cleans GPS pings from about 45 million mobile devices. The data are available daily at census block group resolution and are close to a random sample of the population. Our primary measure of mobility, not social distancing, is the percent of time spent away from home. We calculate this measure based on the median time (in minutes) that a device was observed at its geohash-7 (about 153m x 153m) home location, which SafeGraph determines as the night time residence of the device in the 6-weeks prior. The data cover the entire period of observation from Jan 1, 2019 until the end of April 2020. Demographic Data: We downloaded census block group level demographic information from the U.S. Census Bureau 2018 5-year American Community Survey (ACS) for all CBGs in California using the tidycensus package for the R programming environment (June 29, 2020). Demographic features included ACS sample-based CBG-level estimates of: population count; White race count (alone or in combination with one or more other races), or ``White"; Black or African American race count (alone or in combination with one or more other races), or ``Black"; Asian race count (alone or in combination with one or more other races), or ``Asian"; Hispanic or Latino origin (of any race) count, or ``Hispanic"; and median income. The other census race designations (American Indian or Alaska Native; Native Hawaiian or Other Pacific Islander) represent a substantially lower share of the California population, and were therefore excluded from our analysis due to small sample sizes. The CBG-level ``share" of these groups was calculated by dividing the CBG count by the CBG population. Population density was calculated as the CBG population divided by the area of the CBG. For the aggregate comparison, we compute the share of the non-White population which may be Hispanic as one minus the share of Whites which do not also identify as Hispanic. Because Hispanic is a separate designation from race in the ACS (i.e., those categorized as Hispanic may also be of any race), we evaluated how distinct Hispanic was from race variables of interest. On average, less than 1% of those identified at the CBG level as Hispanic were also identified as Black or Asian; 61% of Hispanic were White. Thus, Hispanic is effectively distinct from Asian and Black categorizations, and we consider Hispanic, Asian, and Black designations to be unique demographic indicators in our model. The baseline reference group in the more detailed comparison contains all other races and ethnicities and therefore consists almost entirely of people who identify as non-Hispanic White.

	Geographic Data: We calculated road density (m/km2) using The Global Roads Inventory Project (GRIP4) vector dataset for North America, downloaded at https://www.globio.info/download-grip-dataset (April 4, 2020). The GRIP4 dataset harmonizes global geospatial datasets on road infrastructure, including road features that can be categorized as highways, primary roads, secondary roads, tertiary roads and local roads. It is consistent with primary and secondary road classifications from the U.S. Census TIGER/Line shapefiles for roads. To calculate road density for each CBG, we summed road lengths within the area of the CBG, and divided by the area of the CBG.
Timing and spatial scale	Analysis covers the period of January 1 - April 30 of both 2019 and 2020 for the state of California. The unit of analysis is the Census Block Group. Gridded satellite observations and weather data are aggregated to the CBG scale for analysis.
Data exclusions	For both CARB and PurpleAir data, days with mean PM2.5 equal to zero or greater than 500 µg/m3 were removed as outliers. Sites for which we removed more than 10% of data were excluded from the entire analysis. Sites with less than 80% data coverage during our study period were also excluded. For our statistical models requiring 2019 and 2020 data, we applied these requirements to both years independently. This quality filtering removed 5.9% of daily CARB PM2.5 data, and 11.4% of daily PurpleAir data, resulting in remaining data from 1664 individual stations (119 CARB and 1545 Purple Air). However, only 830 of those stations (106 CARB and 724 PurpleAir) include data for 2019 and 2020 for the pre-shutdown and shutdown period, and were therefore used in our empirical statistical analysis. This corresponded to 733 unique CBGs with surface PM2.5 measurements. All other data were available for these CBGs.
Reproducibility	Several features of our data and design facilitated robustness checks: The use of two distinct pollutants as outcomes, one measured on the ground and one by satellite, offered one opportunity. The different spatial coverage for the two pollutants, as well as the different spatial coverage of the ground-based monitoring networks also offered another. In addition, we tested many different sensible control and weighting schemes, with very similar results. Complete code is provided for reproducibility.
Randomization	Not applicable. This is a retrospective, quasi-experimental study.
Blinding	Not applicable.

Did the study involve field work? ☐ Yes ☒ No

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 & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging