NBER WORKING PAPER SERIES

# PRIVATE ADOPTION OF PUBLIC GOOD TECHNOLOGIES: THE CASE OF PURPLEAIR

Joshua S. Graff Zivin
Benjamin Krebs
Matthew J. Neidell

Working Paper 32356
http://www.nber.org/papers/w32356

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2024

For comments and suggestions, we thank all participants at the Workshop on Energy Policy and Environmental Economics at Cornell University (June 2023) and the Indoor Air Pollution Workshop at the University of Southern California (August 2023), as well as at research seminars at the Danish Technical University (March 2023) and the MIT Center for Energy and Environmental Policy Research (April 2024). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Private Adoption of Public Good Technologies: The Case of PurpleAir
Joshua S. Graff Zivin, Benjamin Krebs, and Matthew J. Neidell
NBER Working Paper No. 32356
April 2024
JEL No. H41,Q53,Q55

**ABSTRACT**

We study the private adoption and diffusion of a technology that provides a local public good – PurpleAir (PA) pollution monitors. From a purely informational perspective, the ideal spacing of these monitors should reflect the degree of spatial correlation in pollution. In stark contrast, we find that monitor adoption is spatially highly clustered in less polluted areas, suggesting the marginal monitor adopted provides minimal additional public information. Moreover, monitor adoption mainly occurs in affluent, predominantly white neighborhoods, underscoring the potential environmental justice concerns associated with the private provision of this public good.

Joshua S. Graff Zivin
University of California, San Diego
9500 Gilman Drive, MC 0519
La Jolla, CA 92093-0519
and NBER
jgraffzivin@ucsd.edu

Benjamin Krebs
Massachusetts Institute of Technology
Center for Energy and Environmental Policy Research
400 Main Street
Cambridge, Mass 02142-1017
bkrebs@mit.edu

Matthew J. Neidell
Department of Health Policy and Management
Columbia University
722 W 168th Street, 6th Floor
New York, NY 10032
and NBER
mn2191@columbia.edu

# 1 Introduction

The sensor revolution only began in earnest during the past decade, and yet the number of devices that monitor our physical environment and convert that data into a digital signal in real-time is remarkable, providing what the US National Science Foundation calls the world's 'first electronic nervous system' (NSF 2004). While most of these sensors are used by governments, firms, or consumers for purely private purposes, a surprising number also broadcast information in the form of public or quasi-public goods. This includes crowd-sourced data on transportation conditions, criminal activities, as well as weather and other environmental conditions. The Internet of Things and expansion of 6G mobile networks will only serve to increase their prevalence (Imoize et al. 2021; Ang et al. 2022). One interesting feature of the markets for these public-facing sensors is that knowledge spillovers should influence their geographic spread while, at the same time, wealth and other constraints will influence the overall distribution of information across the population. In this paper, we study the adoption and diffusion of one such high-quality sensor — the PurpleAir (PA) monitor — which is sold directly to consumers and measures local air pollution levels.[1]

The PA market is interesting for at least two reasons. First, air pollution is the 4$^{\text{th}}$ largest cause of death globally, leading to 6.7 million premature deaths per year (HEI 2020; Fuller et al. 2022). Since the birth of the environmental movement in the 1960s, most of the regulatory efforts to address this problem have focused on reducing emissions. At the same time, there is an increasing realization that individuals play an important role in their exposure to pollution. In particular, they can engage in temporal avoidance by altering activities on more polluted days, with evidence amassing to indicate that information about air quality leads to changes in behavior (Burke et al. 2022).[2] Doing so effectively, however, requires having real-time pollution data at suitable spatial scales. Regrettably, official government air quality monitoring networks tend to be quite sparse: the US, which has the largest network in the world, has approximately 1 monitor per 750 square miles. In India, that number drops to approximately 1 monitor per 3000 square miles (Guttikunda et al. 2023). Fortunately, the sensor revolution has made inexpensive and reliable air quality monitoring increasingly available, with the promise of measurement at a much finer level of spatial aggregation.[3]

Second, PA produces a high-quality monitor that is sold directly to consumers at an approximate retail cost ranging from $180-$250 US. One particularly interesting feature of this product is that its default setting broadcasts its measures of geocoded pollution through the PA website and app, thereby making that information a local public good. Thus, the extent to which these monitors improve the accuracy of the official monitoring network at fine spatial scales depends on the pattern of monitor adoption. From a purely informational perspective, the ideal spacing of monitors should

[1] The Air Quality Sensor Performance Evaluation Center, maintained by the California South Coast Air Quality Management District, found that PA sensor data was highly correlated with standard reference measures in a controlled laboratory setting (see Krebs et al. (2021) for more details).

[2] In the long-run, wealthier individuals may also sort into less polluted locations (Banzhaf and Walsh 2008).

[3] At the same time, we are seeing tremendous advances in the measurement of pollution using satellite-based data (Van Donkelaar et al. 2021). While that data also provides unprecedented levels of spatial detail, the steps involved in producing it limit the temporal resolution to a monthly scale and cause significant lags in the release of data.

reflect the degree of spatial correlation in pollution. Other factors that may also influence private adoption include income, preferences, and the perceived benefits from hyper-localized air quality measurement. The goal of this paper is to characterize the diffusion of monitors over time and to assess the degree to which that pattern aligns with one that is focused exclusively on a value-of-information approach.[4]

Our paper reveals several interesting patterns regarding monitor adoption. First, we find high spatial correlations between the readings of monitors placed outdoors within the same census tract. Despite these correlations, we observe high spatial clustering of monitors: 19.5% of census tracts have at least 1 outdoor monitor per square mile, and 5.2% of census tracts have over 5 outdoor monitors per square mile, resulting in high levels of information redundancy. This suggests that the marginal outdoor monitor provides limited additional information. In stark contrast, we find low correlations between readings of monitors placed indoors, which is sensible given the contributions of household-specific sources to indoor pollution concentrations. This suggests that the marginal indoor monitor provides high levels of information, but largely as a private value to the homeowner, making the high levels of spatial clustering that we observe more rationalizable in this private good setting.

To better understand the factors behind this clustering, we examine census tract characteristics correlated with monitor adoption, and uncover several interesting findings. As with many new devices, we find that monitor purchases are increasing in income. This finding that monitors are a normal good alone is unsurprising but implies that the placement of monitors may actually increase health inequalities — if monitors enable people to improve their health through avoidance behavior, then the benefit from these monitors are more likely to accrue to the higher income individuals that adopted them. This logic also underlies another more interesting pattern: monitors are more present in less polluted areas, presumably due to residential sorting driven by income. This pattern of adoption reduces the social value from these monitors, since the places that would benefit the most from information that could encourage pollution avoidance behavior are precisely the ones least likely to have this information. We also find that monitor adoption is lowest in areas with a higher share of Black or Hispanic populations, suggesting the monitors create racial inequality in access to pollution information as well.

We delve further into these findings by exploring dynamic patterns of monitor adoption, first exploring how changes in pollution relate to monitor purchases. Although our cross-sectional results indicate that monitors are more present in less polluted areas, we find that increases in pollution within an area — notably from wildfires — lead to increases in monitor purchases. These findings are consistent with Bayesian learning, in which households update their beliefs about the value of a monitor as new pollution shocks occur.

[4]Our analysis mirrors the extensive literature on technology adoption and diffusion which began with Zvi Griliches's seminal work on hybrid corn (Griliches 1957) and has been extended to examine the modern technologies of our digital economy (e.g., Caselli and Coleman 2001) with one important distinction. Our analysis focuses on the adoption of a technology that provides public goods, which greatly complicates the spatial elements of diffusion relative to this prior literature that has examined goods whose value is largely private. See Hall (2004) for a review of the core issues and challenges.

In the dynamic specification we also explore peer effects by including the lagged cumulative number of monitors within a tract as a control variable. We find that past monitor purchases by peers within a tract is highly predictive of current purchases. Such results may reflect true peer effects, whereby individuals learn about these monitors from their neighbors, or unobserved correlates that drive the monitor purchases of all households within a neighborhood, such as a taste for new technology.

Given the low marginal informational value of each additional outdoor monitor in a nearby location, we explore several possible drivers behind the strong patterns of clustering. We largely rule out explanations based on incomplete information. The process behind monitor purchase decisions, where customers must visit the PA website, makes it unlikely that consumers are unaware of the correlation between proximate monitors. Consistent with this, we find that additional monitors placed within a tract do not improve measures of pollution variability within the tract. Of course, a high correlation on average does not necessarily mean that monitors are highly correlated at all points in the pollution distribution, suggesting that the clustering of monitors may be justified based on an option value approach.

We extend this exercise to explore the correlation of monitor readings along various points in the pollution distribution but find minimal evidence to support an option value explanation. The residual explanation for this pattern of monitor clustering relies on spatial clustering in preferences, although we cannot test these explanations directly. Technophiles may purchase monitors for reasons that are quasi-independent from the value of information that the monitor provides, including a competitive desire to 'keep up with the Joneses' and thus drive high levels of spatial correlation in monitor adoption.

Although the sensor revolution holds promise to greatly improve lives, our results suggest that a market-based allocation leads to suboptimal information provision. We recognize the challenge in inferring welfare without knowing the full value from monitor adoption, and revealed preferences suggest consumers are better off purchasing them. Nonetheless, an optimal policy based on a value of information approach will require supplemental provision of monitors where the private market falls short. The efficiency of any such policy will, of course, depend on design features that limit the crowding out of private purchases, underscoring the need to better understand the drivers of private market adoption in the first place.

These shortcomings of market-based information provision also have implications for environmental justice. Given the spatial variation in pollution levels, publicly provided air quality monitors provide a local public good in which the informational value of a monitor declines with distance. Privately provided monitors, which supplement the publicly provided monitors, provide the most value in the areas where they are placed. Since their placement occurs in wealthier, Whiter, and cleaner areas, the private market for monitors leads to inequality in access to information, a pattern that could be alleviated through coordinated monitor placement.

## 2 Background and Data

We focus our analysis on ambient particulate matter less than 2.5 microns in diameter ($PM_{2.5}$), small particles that have severe negative health consequences when inhaled (EPA 2023b). $PM_{2.5}$ comes from both outdoor and indoor sources. Outdoor sources include agricultural processes, industrial combustion, power plants, and transportation, but also non-anthropogenic sources, such as wildfire smoke (McDuffie et al. 2021). In the indoor environment, $PM_{2.5}$ originates from outdoor sources that infiltrate indoors[5], as well as from indoor sources such as cooking (EPA 2023a).[6]

Information about outdoor vs. indoor pollution has distinct implications for short-run adaptive behavior. The level of outdoor $PM_{2.5}$ in a given location is not controllable by the individual. Thus, avoidance must come from changing the type or location of one's activities, such as staying inside or spending time in less polluted areas (e.g., Neidell 2009; Burke et al. 2022). Indoors, however, people can control the level of pollution through various behavioral responses. They can alter their contribution to indoor emissions through cooking decisions or reduce emissions by running air purifiers or opening windows (as long as outdoor levels are lower than indoor levels) (Metcalfe and Roth 2024).

Measures of outdoor and indoor monitor adoption come from the PurpleAir Real-Time Air Quality Monitoring Network. PA monitors report $PM_{2.5}$ estimates from laser-based particle counts. The monitors are not approved by the EPA for regulatory purposes but have proven to be highly reliable in performance evaluations (AQMD 2019) and have recently been adopted by the EPA as an informational tool (EPA 2022).[7] PA monitors lack a display and must be connected to the internet for the owner to obtain $PM_{2.5}$ readings.[8] Once installed, the monitors upload their readings to the PA website, where the purchaser and anyone else can freely access the monitor readings, unless the owner actively opts-out of the default public setting.

We downloaded daily readings[9] of California-based PA monitors for the years 2019-2021 from both outdoor and indoor monitors.[10] This time period includes major wildfire events like the August Complex fire in 2020, which allows us to assess the impact of sustained elevated $PM_{2.5}$ levels on monitor adoption. The GPS coordinates of the monitor allow us to identify the census tract in which the monitor resides[11], enabling us to measure the placement of monitors relative to each other and to merge other spatial information. Indeed, PA monitors have proliferated extensively in

[5] Using PA data, Krebs et al. (2021) show that outdoor $PM_{2.5}$ enters the indoor environment quickly and to a large extent.

[6] Cooking contributes to $PM_{2.5}$ not through the kind of stove used but through the foods cooked and types of cookware used.

[7] Data from PA monitors are now available as a layer in Google Maps.

[8] A more recent monitor design provides a display of pollution levels, but this feature became available after our analysis period.

[9] $PM_{2.5}$ readings are reported at 2-minute intervals, but PA provides pre-aggregated data.

[10] When customers install their PA monitor, they self-report whether it is installed outdoors or indoors. On the PA website, indoor monitors are denoted by a bold ring around the pollution readings.

[11] Purchasers can elect to toggle their exact location randomly by 500 feet, then the random location becomes fixed at the time of installation. This introduces a classical measurement error that reduces precision in all estimates and introduces attenuation for the dynamic estimates.

California, increasing from 947 outdoor and 247 indoor monitors in January 2019 to 7506 outdoor and 3876 indoor monitors in December 2021, a nearly 10-fold increase. This growth is illustrated in Appendix Figure A1, which also highlights the monitors' uneven distribution across the state, with higher concentrations in major metropolitan areas.

We merge three other sources of data to the PA data. Using the census tract where the monitor is placed, we first merge various demographic variables from the 2019 American Community Survey. Since monitors are endogenously located, making pollution measures a select sample, we also merge satellite-derived data on $PM_{2.5}$ concentrations from Van Donkelaar et al. (2021), which provides measures of monthly data at the $0.01° \times 0.01°$ level for the entire globe, corresponding to less than one square mile in California. We merge this to the census tract by calculating the weighted average of overlapping grid cells, where the weights reflect the share of the tract coinciding with a grid cell. Using the exact coordinates of official EPA air quality monitors, we calculate the distance from the geographic centroid of the census tract to the closest EPA $PM_{2.5}$ monitor. Table A1 reports the means and standard deviations of these variables for all census tracts in California. As might be expected with a new technology, those tracts with at least one monitor are different than those without any, which we explore directly in our estimation of the drivers of monitor adoption.

To highlight the additional information provided by the placement of a new monitor, Figure 1 presents a scatter plot of the census tract-level average $PM_{2.5}$ concentrations in 2021 for monitors operational in January 2019 vs. monitors added to the same tract in between February 2019 and December 2020, separately for outdoor (a) and indoor (b) monitors. The plot reveals distinct correlations between outdoor and indoor monitors, with a correlation of 0.85 between proximate outdoor monitors and 0.22 for proximate indoor ones. These differences yield two important points regarding the marginal value of outdoor vs. indoor monitors. First, since measurements from outdoor monitors close to each other are highly correlated, the additional public informational value of a new outdoor monitor close to existing monitors is low, implying information is increasing in the distance from the nearest monitor. Second, since indoor monitors close to each other are not highly correlated, the additional private informational value from a new indoor monitor is high regardless of its proximity to surrounding monitors. This is not surprising because of the role of indoor sources in contributing to pollution levels. Therefore, from a purely informational perspective, the clustering of indoor monitors provides high private value, whereas the clustering of outdoor monitors provides low value.

## 3 Econometric model

Our empirical focus is to estimate the influence of several factors on outdoor or indoor monitor adoption. Several features guide our specification. Given the differential value of outdoor vs. indoor monitors, we estimate models separately for each. Since the price of PA monitors has been relatively stable over time, we assume a constant supply and specify a single equation model. We treat the number of monitors within a tract as count data and estimate a Poisson pseudo-maximum likelihood regression model, which relaxes the assumption that the mean equals the variance (Gourieroux et al. 1984). We report the exponentiated coefficients from this regression, which reflect the incidence rate ratio, then subtract 1 from these coefficients to reflect the percentage change in monitor prevalence from a one-unit change in an independent variable.

We begin by estimating a static, cross-sectional model according to the following equation:

$$M_c^j = exp(X_c' \delta_1^j + P_c' \beta_1^j + S_c' \gamma_1^j + \tilde{X}_c' \tilde{\delta}_1^j + \varepsilon_c^j) \tag{1}$$

$M_c^j$ refers to the total number of monitors in census tract $c$ by the end of 2021 in $j = \{\text{outdoor}, \text{indoor}\}$ setting. $X_c$ contains a subset of the census tract demographic of particular interest, described in more detail below. The vector $P_c$ includes two satellite-derived pollution measures: the average and the standard deviation of $\text{PM}_{2.5}$ concentrations over the 2019-2021 period. $S_c$ is the distance from the census tract centroid to the closest EPA monitor in miles. $\varepsilon_c$ denotes the error term.

Our analysis concentrates on several variables as potential drivers of monitor adoption. Since monitor adoption is likely a normal good, we focus on household income as a predictor of monitor demand. Education serves as a proxy for consumer awareness of this relatively new product, though its impacts may differ for indoor and outdoor monitors given their different informational value. Local pollution levels also likely influence monitor adoption, where the value of monitors is higher in more polluted areas. To explore potential environmental justice angles, we also explore the role of Black and Hispanic population shares in monitor adoption.

This analysis is a descriptive one. We recognize that many factors likely confound the relationship between these characteristics and monitor adoption. For example, individuals with higher health risks from pollution exposure may be more likely to live in cleaner areas but also more likely to purchase monitors. Given the cross-sectional nature of our analysis, we cannot definitively rule out these concerns. We nonetheless enhance this model by including the remaining census tract variables from Table A1 in the vector $\tilde{X}_c$, and probe the sensitivity of estimates to their inclusion.

The analysis of race and ethnicity, however, is a notable exception, since these characteristics are likely be influenced by the economic variables included in our regression models. For example, discrimination in the labor market may lead to lower wages for certain racial and ethnic minorities, such that income may be a potential mediator affecting monitor adoption. As a result, models that include both race and income are difficult to interpret because income would constitute a 'bad control' (Angrist and Pischke 2009) that leads to an overcontrol bias (Cinelli et al. 2022). To address this issue, we present models that include race and ethnicity in addition to the other variables, but we also estimate models that only include race and ethnicity to obtain estimates of the composite impact of these characteristics. This composite impact reflects the effects of race directly as well as the impacts of race on income and education, due to systemic injustices, on adoption decisions. Both measures shed important light on potential environmental justice concerns related to the private provision of this local public good.

We further extend our static analysis to explore dynamic patterns of monitor adoption, estimating the following equation:

$$M^j_{ct} = exp([\sum_{k=0}^{t-1} M'_{ck}]\alpha^j_2 + L(P'_{ct})\lambda^j_2 + X'_c\delta^j_2 + P'_c\beta^j_2 + S'_c\gamma^j_2 + \tilde{X}'_c\tilde{\delta}^j_2 + \theta^j_t + \varepsilon^j_{ct}) \tag{2}$$

$M^j_{ct}$ denotes the number of new monitors added in tract $c$ in month $t$ in $j = \{\text{outdoor}, \text{indoor}\}$ setting. The summation term measures the cumulative monitor base in the past month, i.e., the total number of monitors online in the previous month. By including the cumulative number of monitors, we examine the influence of prior monitor adoption by neighbors on current monitor adoption. As more households in a given tract purchase monitors, this may increase the awareness and appeal of the monitor, leading to a potential peer effect.

We include separate measures of cumulative outdoor and indoor monitors in each equation to explore if own-product and cross-product prior adoption have differential impacts. Although we establish temporal precedence between past and new monitor adoption, peer effects are likely to be endogenous and hence challenging to identify (Manski 1993). For example, unobserved factors within a tract, such as a preference for new gadgets, may lead to more monitor purchases both in the past and currently. As a result, we interpret the coefficient of $\alpha_2$ as an endogenous peer effect that captures both the exogenous peer effect component and the impact of unobserved factors common to peers.

We control for changing pollution levels within a tract in the term $L(P_{ct})$, which we specify as the current and past month average $PM_{2.5}$ levels separately. These terms allow us to explore potential learning about pollution levels. Although long-term pollution levels can influence monitor adoption, short-term changes in pollution levels may lead consumers to update their beliefs about pollution or increase the salience of pollution concerns, thus influencing demand (Burke et al. 2022). For example, spikes in pollution driven by high-pollution events, such as wildfires, may lead to a surge in monitor demand. Our controls for pollution levels in the same and previous month enable us to explore whether people updated their beliefs about pollution levels based on these new experiences.

We continue to control for the same factors as in the static equation, though their interpretation changes considerably since the dependent variable is the number of new monitors added in a given month. We also include year-season dummy variables ($\theta_t$) to account for time trends, and we cluster regressions on the census tract to account for arbitrary serial correlation within tracts.

## 4 Results

We begin with visual evidence for monitor clustering by mapping active outdoor and indoor pollution monitors, zooming in on the San Francisco Bay Area.[12] Figure 2 shows monitor placement as of December 2021, along with several tract-level measures. Panel (a) shows that monitors, while

[12] We also calculate Moran's I, a measure of spatial autocorrelation, at the monthly level for outdoor and indoor monitors in the entire state. Over the three-year time period, Moran's I increased from 0.20 to 0.49 for outdoor monitors and from 0.12 to 0.48 for indoor monitors, indicating high levels of clustering.

highly correlated, are more prevalent in some areas than in others, with San Francisco and Berkeley clearly having higher monitor densities than San Jose and Fremont. As Panel (b) shows, monitors cluster in places with high income, suggesting that monitors are, unsurprisingly, a normal good. This clustering by income is potentially concerning. Panels (c) and (d) demonstrate that monitor adoption is denser in areas with relatively good air quality and higher White population shares, respectively, suggesting potential inequality concerns.

Our static regression results, shown in Table 1, build on the visual evidence just discussed by expanding our focus to the entire state and adjusting for additional control variables. Recall that we provide the incidence-rate ratio from the Poisson model in Equation 1, where subtracting 1 yields the percentage change in the dependent variable from a one-unit change in the explanatory variable. Columns 1-3 refer to outdoor monitors and Columns 4-6 refer to indoor monitors. Columns 2 and 4 include the additional demographic controls in $\tilde{X}_c$ to assess the sensitivity of our estimates, while Columns 3 and 6 isolate the race and ethnicity variables.

As expected, we find that tracts with higher incomes have more monitors. This pattern is true for both outdoor and indoor monitors, with estimates of comparable magnitude across the two models. A \$10,000 increase in household income correlates with a statistically significant 4% increase in the number of monitors in the tract. The estimates for income are quite stable to the addition of further demographic controls.

Monitor adoption also increases with education levels. As the share of the population with a bachelor's degree increases by 0.1, outdoor monitor adoption increases by 28%. The estimate for indoor monitors implies an astonishing rise of 712%. The difference in magnitude between the outdoor and indoor equations could reflect understanding about the difference in the marginal value of a monitor, whereby indoor monitors provide high value regardless of the presence of nearby monitors. This difference in value from monitor placement may not be readily understood by many people, and additional schooling could be an important factor in understanding this difference. These estimates are sensitive to adding more demographic controls, though they remain statistically significant and in the same direction.

Turning to the results by race and ethnicity, we first focus on the unadjusted models, where the coefficients reflect the composite impact of race and ethnicity on monitor adoption described earlier. These results, shown in Columns 3 and 6, indicate a large impact of race and ethnicity. As the Black population share increases by 0.1, we estimate a 10% decrease in outdoor monitors and an 8% decrease in indoor monitors. Similar estimates for the Hispanic population share indicate a 10% decrease for both outdoor and indoor monitors. As we add additional control variables, shown in Columns 2 and 5, the coefficients on race change considerably and with different patterns across monitor placement. For outdoor monitors, estimates continue to show a decrease in monitor purchases for both Black and Hispanic populations, though smaller in magnitude than the unadjusted estimates. For indoor monitors, we do not find a statistically significant estimate for more Hispanic areas, and we find an increase in purchases for areas with more Black residents. As previously discussed, many of these control variables may lead to overcontrol bias by capturing the mediating path by which race and ethnicity affects monitor adoption, so we interpret these adjusted results with caution.

Focusing on pollution levels, we find that cleaner areas have significantly higher rates of monitor adoption. A 1 µg/m$^3$ lower level of PM$_{2.5}$ increases outdoor monitors by 33% and increases indoor monitors by 36%. This result is striking. In areas where pollution is the highest, and thus avoidance behaviors are potentially the most effective, people have less knowledge of their pollution levels, even when conditioning on income and education. Perhaps more economically rational, areas with more variation in pollution have more monitors. A one-standard-deviation change in PM$_{2.5}$ leads to 17% and 23% increases in outdoor and indoor monitors, respectively, suggesting that variation in pollution is an important factor in driving people to learn about pollution levels. The closer the census tract is to an official EPA monitor, the fewer monitors are present in that tract. Conditional on our full set of controls, a census tract that is 1 mile closer to an EPA monitor has 0.8% fewer outdoor monitors, consistent with the notion that PA monitors are a substitute for EPA monitors. We also find that indoor monitor adoption falls in census tracts closer to an EPA monitor, suggesting that consumers also view them as substitutes — perhaps because the salience of the EPA monitor makes a personal one, even if placed indoors, feel less valuable.

We next turn to the dynamic results, displayed in Table 2, first focusing on endogenous peer effects. For own-product impacts, we find that monthly monitor purchases increase with the cumulative number of monitors within that tract. One additional outdoor monitor purchased in the past leads to a 11% increase in current outdoor monitor purchases within a census tract. The comparable estimate for indoor monitors is 6%. This endogenous peer effect underpins the spatial clustering shown in previous figures.

In terms of cross-product impacts, we obtain different results for outdoor and indoor monitors. Current outdoor monitor adoption is unrelated to past indoor monitor purchases. The estimates are statistically insignificant and, at less than one percent, small in magnitude. Current indoor purchases, however, are strongly related to past outdoor monitor purchases, indicating an 8% increase from one additional outdoor monitor. Since placing monitors indoors produces high private value, we speculate that the differential result by monitor placement indicates that people learn about monitors regardless of where they are placed, but that they correctly infer the high private value from a monitor placed indoors and the low private (and public) value from placing a monitor outdoors.

We next focus on contemporaneous and lagged PM$_{2.5}$ to explore potential learning. As PM$_{2.5}$ levels increase, we find statistically significant increases in both outdoor and indoor monitor adoption, with a larger impact on indoor monitor adoption (2.2% vs. 1.1% increase). Furthermore, the impact from the current month is larger than the past month for both outdoor and indoor monitors. These results are consistent with people learning about pollution levels over time, leading to an increase in monitor purchases.

Taking stock of our results, the adoption pattern for indoor monitors is consistent with an economic model in which adoption decisions are driven by information value, whereas the pattern for outdoor monitors is somewhat puzzling given their low private value. What might explain the clustering of outdoor monitors?

One possible explanation is incomplete information, whereby people are unaware of the high spatial correlation in outdoor pollution and thus the monitor's low marginal value. While this is plausible for early adopters, it seems unlikely beyond the first handful of monitors within a tract. Purchasing a monitor requires consumers to visit the PA website, where they are confronted with data showing all the monitor readings near their home. If a household member sees that measures are highly correlated within their tract, this should make clear to them that purchasing another outdoor monitor will provide limited additional information. As a test of this, we regress the $PM_{2.5}$ measures from the first monitor on subsequent monitors, sequentially adding each new monitor to the regression.[13] The black dots in Figure 3 display the absolute $R^2$ (top panel) and the change in $R^2$ (bottom panel) as more monitors are added in a tract. The $R^2$ hovers around 0.9 regardless of the number of monitors, scarcely changing as more monitors are added to a tract. This confirms that additional monitors provide minimal information, thus making incomplete information an unlikely explanation for our main results.

Another potential driver of this pattern of adoption is option value. Our results thus far have shown that geographically proximate monitors are highly correlated on average, but adopters may be more concerned about correlation at particular points in the pollution distribution. In particular, we might imagine that consumers are most concerned with high pollution days, and it may well be that hyper-localized information is more valuable in those cases. To assess the plausibility of such an explanation, we extend the previous analysis to measure the same $R^2$ but focus on pollution readings in the $75^{\text{th}}$ and $90^{\text{th}}$ percentile (instead of the mean) of the pollution distribution. These results, also shown in Figure 3, show somewhat lower $R^2$ levels for the first few monitors compared to the mean, though the values are still quite high. By the fourth monitor, however, the differences between these percentiles and the mean disappear. This suggests that the option value from additional monitors, at least beyond the first few in a locality, is quite low.

Other possible explanations for the clustering of outdoor monitors rely on non-random sorting of individuals into neighborhoods that leads to spatial clustering in preferences. For example, technophiles — individuals who have a penchant to adopt all kinds of new technologies — may choose to live near each other and also be more likely to buy new gadgets that measure air quality quasi-independently from the value of information provided by that monitor. A desire to compete with like-minded neighbors could also generate high levels of spatial correlation because others see the broadcasted pollution of their neighbors or because they brag about their shiny new toy at dinner parties, kids' soccer games, or the country club. Regardless, the utility generated from monitor adoption, in this case, is generated through ownership rather than the information services provided by that ownership. Unfortunately, direct tests of preference clustering of this sort are

[13] To construct this dataset, we used monitors that had less than 25% missing observations in 2021 to remove intermittent monitors that people may not necessarily use as a reliable source of information. We include census tracts with at least 10 operational monitors. We run daily-level linear regressions for 2021 with tract, month, and day-of-week fixed effects.

exceedingly difficult to construct with the limited details available about consumers. As such, we offer this as a plausible explanation for the clustering patterns that cannot be explained by the previous two channels. Unpacking the mechanisms driving this 'residual' explanation is an area rich for future work.

## 5 Conclusion

Residential-grade air quality monitors are a highly sought-after product, clearly valued by consumers. While these monitors provide considerable new information to households when placed indoors, their value outdoors depends on their proximity to other nearby monitors. In this paper, we document a pattern of significant spatial clustering in the adoption of outdoor monitors that limits the informational content of the monitoring network. This suboptimal pattern of adoption does not appear to be driven by incomplete information or option value. Instead, we conclude that some unobserved and spatially clustered preferences, such as a desire for new technology or an effort to keep up with one's neighbors, is likely driving the adoption and diffusion of this novel technology.

In this context, our results highlight one of the core challenges arising from private provision of a public good. The preferences of consumers are unlikely to fully align with social objectives since the efficient production of additional information from monitors requires them to be evenly spread across the population. Not only does clustering yield less informational content than it might otherwise do for any given level of monitors, but it also contributes to inequality in access to information about air quality. This inefficient and inequitable outcome resulting from uncoordinated private adoption suggests an important potential role for supplementary investments by the public sector. The optimal design of such a government program should leverage the private value consumers appear to receive from ownership (rather than information) to minimize moral hazard and crowd out from these public investments. At the same time, the value of these public investments is likely to extend beyond that experienced by consumers, as it would facilitate research on a less selected sample than those based on the PurpleAir network alone.

An important caveat to these conclusions is that the results are context-specific. Our study is in a high-income country with a high level of existing information about air pollution. The marginal value of monitors is likely much higher in a low-information setting, especially when no other monitors exist. Roughly one-third of countries around the globe lack legal requirements for air quality monitoring (UNEP 2022), suggesting potentially high value from the private adoption of consumer monitors. Nonetheless, as monitors expand into these areas, our analysis hints at the importance of coordinating that expansion.

## References


Ang, Kenneth Li Minn, Jasmine Kah Phooi Seng, and Ericmoore Ngharamike (2022). Towards crowdsourcing internet of things (crowd-iot): Architectures, security and applications. *Future Internet 14*(2), 49.

Angrist, Joshua D and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, New Jersey: Princeton University Press.

AQMD, South Coast Air Quality Management District Air Quality Sensor Performance Evaluation Center (2019). Field evaluation: Purple Air (PM-II) PM Sensor. Available at: www.aqmd.gov/docs/default-source/aq-spec/field-evaluations/purple-air-pa-ii—field-evaluation.pdf?sfvrsn=4, accessed March 25, 2024.

Banzhaf, H Spencer and Randall P Walsh (2008). Do people vote with their feet? An empirical test of Tiebout's mechanism. *American Economic Review 98*(3), 843–863.

Burke, Marshall, Sam Heft-Neal, Jessica Li, Anne Driscoll, Patrick Baylis, Matthieu Stigler, Joakim A Weill, Jennifer A Burney, Jeff Wen, Marissa L Childs, and Carlos F Gould (2022). Exposures and behavioural responses to wildfire smoke. *Nature Human Behaviour 6*(10), 1351–1361.

Caselli, Francesco and Wilbur John Coleman (2001). Cross-country technology diffusion: The case of computers. *American Economic Review 91*(2), 328–335.

Cinelli, Carlos, Andrew Forney, and Judea Pearl (2022). A crash course in good and bad controls. *Sociological Methods & Research*.

EPA, United States Environmental Protection Agency (2022). EPA research improves air quality information for the public on the AirNow Fire and Smoke Map. Available at: https://www.epa.gov/sciencematters/epa-research-improves-air-quality-information-public-airnow-fire-and-smoke-map, accessed March 25, 2024.

EPA, United States Environmental Protection Agency (2023a). Indoor air quality. Available at: https://www.epa.gov/report-environment/indoor-air-quality, accessed March 25, 2024.

EPA, United States Environmental Protection Agency (2023b). Research on health effects from air pollution. Available at: https://www.epa.gov/air-research/research-health-effects-air-pollution, accessed March 25, 2024.

Fuller, Richard, Philip J Landrigan, Kalpana Balakrishnan, Glynda Bathan, Stephan Bose-O'Reilly, Michael Brauer, Jack Caravanos, Tom Chiles, Aaron Cohen, Lilian Corra, Maureen Cropper, Ferraro Greg, Jill Hanna, David Hanrahan, Howard Hu, David Hunter, Gloria Janata, Rachael Kupka, Bruce Lanphear, Maureen Lichtveld, Keith Martin, Adetoun Mustapha, Ernesto Sanchez-Triana, Karti Sandilya, Laura Schaefli, Joseph Shaw, Jessica Seddon, William Suk, Martha María Téllez-Rojo, and Chonghuai Yan (2022). Pollution and health: A progress update. *The Lancet Planetary Health 6*(6), e535–e547.

Gourieroux, Christian, Alain Monfort, and Alain Trognon (1984). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica*, 701–720.

Griliches, Zvi (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 501–522.

Guttikunda, Sarath, Nishadh Ka, Tanushree Ganguly, and Puja Jawahar (2023). Plugging the ambient air monitoring gaps in India's national clean air programme (NCAP) airsheds. *Atmospheric Environment 301*, 119712.

Hall, Bronwyn H (2004). Innovation and diffusion. In J. Fagerberg, D. C. Mowery, and R. R. Nelson (Eds.), *Handbook of Innovation*, pp. 459–485. Oxford, UK: Oxford University Press.

HEI, Health Effects Institute (2020). State of global air 2020. Special report. Available at: https://www.healthdata.org/research-analysis/library/state-global-air-2020, accessed March 25, 2024.

Imoize, Agbotiname Lucky, Oluwadara Adedeji, Nistha Tandiya, and Sachin Shetty (2021). 6G enabled smart infrastructure for sustainable society: Opportunities, challenges, and research roadmap. *Sensors 21*(5), 1709.

Krebs, Benjamin, Jennifer Burney, Joshua Graff Zivin, and Matthew Neidell (2021). Using crowd-sourced data to assess the temporal and spatial relationship between indoor and outdoor particulate matter. *Environmental Science & Technology 55*(9), 6107–6115.

Manski, Charles F (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies 60*(3), 531–542.

McDuffie, Erin E, Randall V Martin, Joseph V Spadaro, Richard Burnett, Steven J Smith, Patrick O'Rourke, Melanie S Hammer, Aaron van Donkelaar, Liam Bindle, Viral Shah, Lyatt Jaeglé, Gan Luo, Fangqun Yu, Jamiu A Adeniran, Jintai Lin, and Michael Brauer (2021). Source sector and fuel contributions to ambient $PM_{2.5}$ and attributable mortality across multiple spatial scales. *Nature Communications 12*(1), 1–12.

Metcalfe, Robert and Sefi Roth (2024). Making the invisible visible: The impact of real-time feedback on indoor air pollution. *Unpublished manuscript*.

Neidell, Matthew (2009). Information, avoidance behavior, and health: The effect of ozone on asthma hospitalizations. *Journal of Human Resources 44*(2), 450–478.

NSF, National Science Foundation (2004). The sensor revolution. A special report. Available at: https://www.nsf.gov/news/special_reports/sensor/SensorRevolutionNSF.pdf, accessed March 25, 2024.

UNEP, United Nations Environment Programme (2022). Annual report 2021. Available at: https://wedocs.unep.org/bitstream/handle/20.500.11822/37946/UNEP_AR2021.pdf, accessed March 25, 2024.

Van Donkelaar, Aaron, Melanie S Hammer, Liam Bindle, Michael Brauer, Jeffery R Brook, Michael J Garay, N Christina Hsu, Olga V Kalashnikova, Ralph A Kahn, Colin Lee, Robert C Levy, Alexei Lyapustin, Andrew M. Sayer, and Randall V. Martin (2021). Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology 55*(22), 15287–15300.

# Tables & Figures

Table 1: Static regression results

| Dependent variable: Nb. monitors | Outdoor monitors | | | Indoor monitors | | |
|---|---|---|---|---|---|---|
| Med. annual houshold income, 1000 USD | 1.004*** | 1.004*** | | 1.002*** | 1.004*** | |
| | (0.000) | (0.000) | | (0.000) | (0.001) | |
| Share Bachelor's degree | 3.790*** | 6.983*** | | 72.177*** | 66.070*** | |
| | (0.385) | (0.881) | | (10.194) | (11.621) | |
| Share Black population | 0.233*** | 0.876 | 0.010*** | 2.890*** | 4.943*** | 0.176*** |
| | (0.058) | (0.214) | (0.003) | (0.824) | (1.442) | (0.048) |
| Share Hispanic population | 0.215*** | 0.653*** | 0.008*** | 0.630** | 0.993 | 0.001*** |
| | (0.024) | (0.080) | (0.001) | (0.117) | (0.206) | (0.000) |
| Av. $PM_{2.5}$ concentration, µg/m$^3$ | 0.674*** | 0.685*** | | 0.638*** | 0.630*** | |
| | (0.006) | (0.006) | | (0.008) | (0.008) | |
| $PM_{2.5}$ standard deviation, µg/m$^3$ | 1.174*** | 1.180*** | | 1.232*** | 1.250*** | |
| | (0.003) | (0.004) | | (0.006) | (0.008) | |
| Distance to closest EPA monitor, miles | 1.000 | 0.992*** | | 0.950*** | 0.949*** | |
| | (0.001) | (0.002) | | (0.003) | (0.004) | |
| *Control variables* | | | | | | |
| $\tilde{X}_c$ | No | Yes | No | No | Yes | No |
| Census tracts | 7,965 | 7,965 | 7,965 | 7,965 | 7,965 | 7,965 |

Notes: Coefficients and standard errors (in parentheses) of the Poisson pseudo-maximum likelihood regressions of the census tract-level number of outdoor monitors (Columns 1-3) and indoor monitors (Columns 4-6) on the listed variables. Columns 2 and 5 further include $\tilde{X}_c$ as control variables. The number of monitors refers to the number of active monitors in December 2021. Average $PM_{2.5}$ concentration and standard deviation covers the years 2019-2021. All demographic variables are from the 2019 American Community Survey.

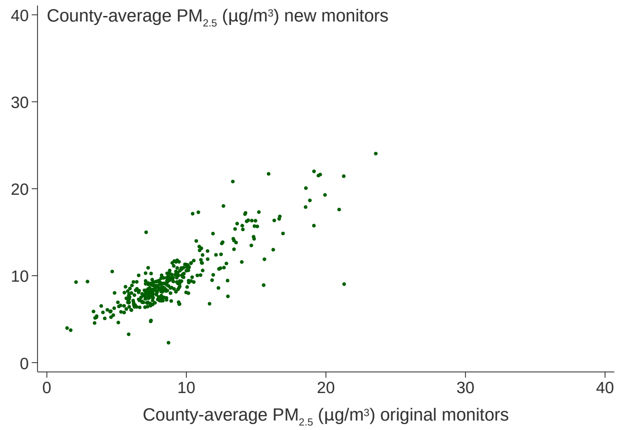
Table 2: Dynamic regression results: lagged $PM_{2.5}$

| Dependent variable: Nb. new monitors | Outdoor monitors | Indoor monitors |
|---|---|---|
| Nb. outdoor monitors past month | 1.111*** | 1.078*** |
| | (0.009) | (0.016) |
| Nb. indoor monitors past month | 0.991 | 1.061*** |
| | (0.023) | (0.010) |
| $PM_{2.5}$ level same month, µg/m$^3$ | 1.011*** | 1.022*** |
| | (0.001) | (0.002) |
| $PM_{2.5}$ level past month, µg/m$^3$ | 1.009*** | 1.009*** |
| | (0.001) | (0.002) |
| *Control variables* | | |
| $X_c$ | Yes | Yes |
| $S_c$ | Yes | Yes |
| $\tilde{X}_c$ | Yes | Yes |
| $\theta_t$ | Yes | Yes |
| Census tracts | 7,965 | 7,965 |
| Observations | 278,775 | 278,775 |

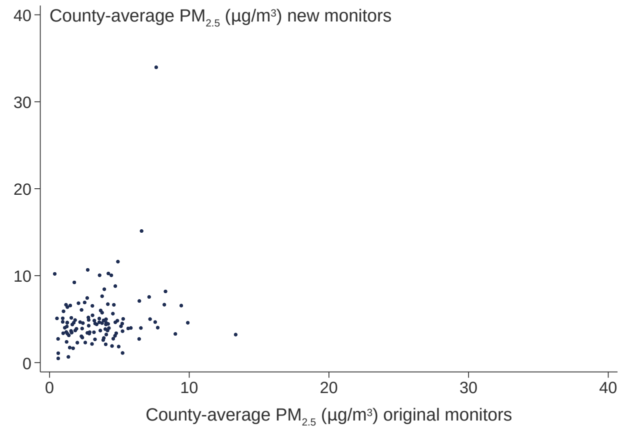
Notes: Coefficients and standard errors (in parentheses) of the Poisson pseudo-maximum likelihood regressions of the monthly census tract-level number of new outdoor (Column 1) and new indoor monitors (Column 2) on the cumulative number of previously added outdoor and indoor monitors and the same- and lagged-monthly average $PM_{2.5}$ levels. Both regressions further include $X_c$, $S_c$, $\tilde{X}_c$, and $\theta_t$ (season, year, and season-times-year indicators) as control variables. The standard errors are clustered on census tracts.

Figure 1: Correlation of monitor readings within the same census tract

(a) Outdoor monitors

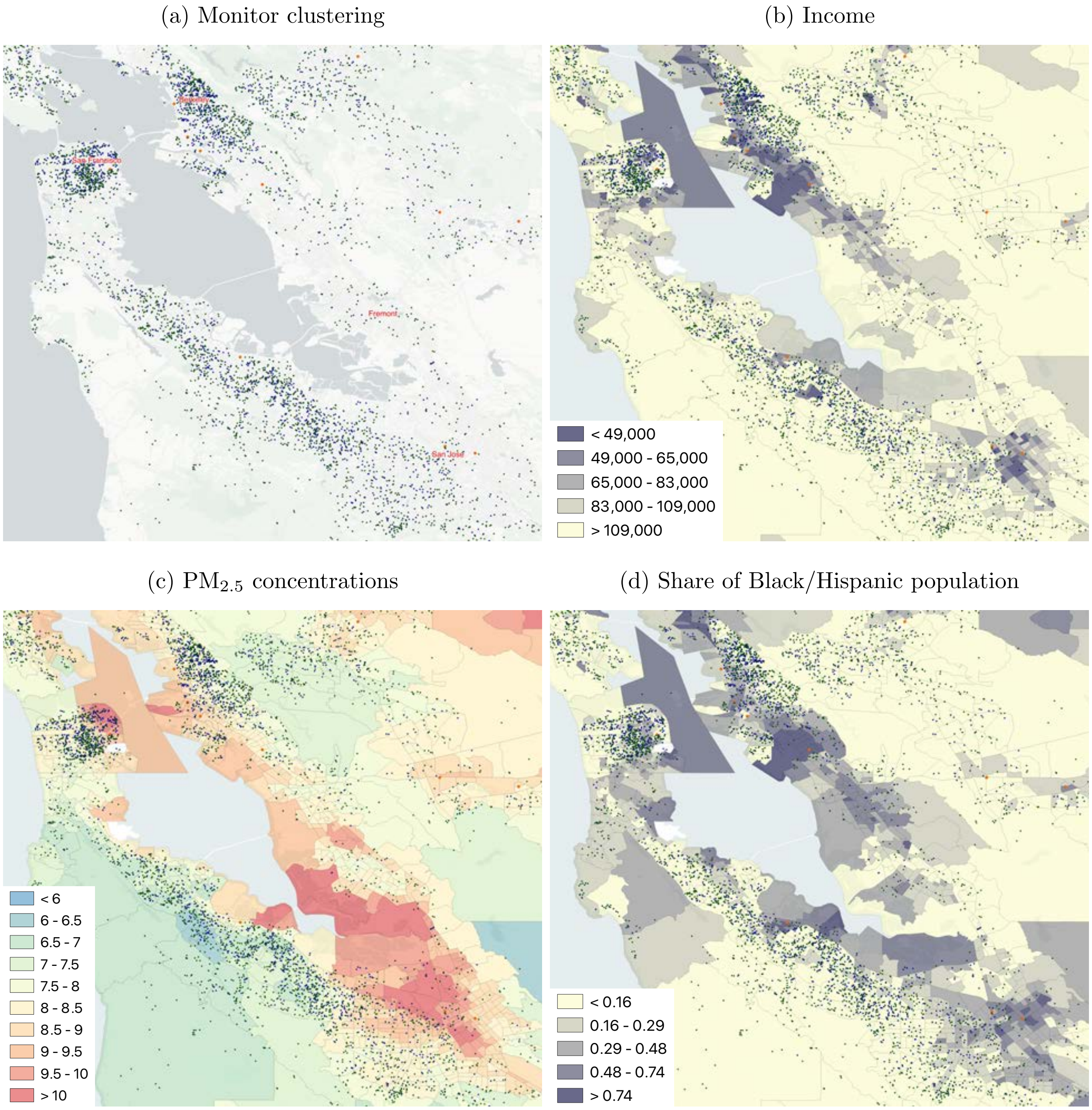



(b) Indoor monitors



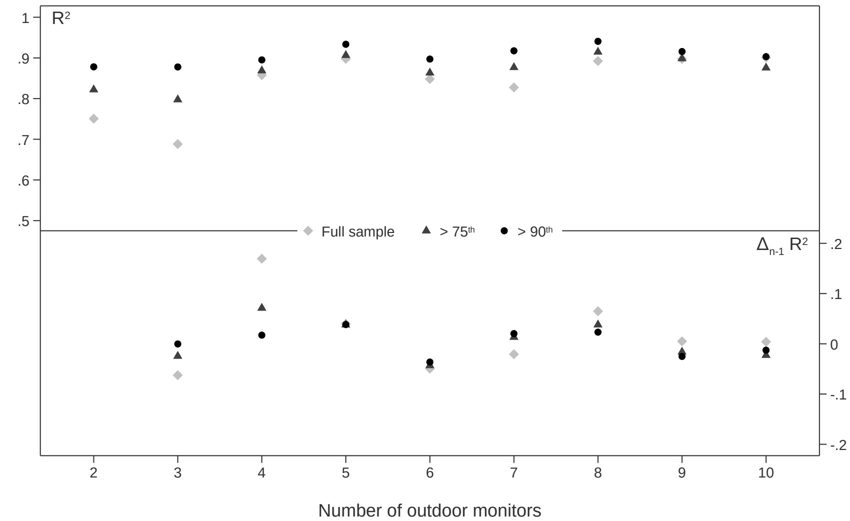

Notes: The figure compares the 2021 county-level average $PM_{2.5}$ concentrations of same-census tract monitors that were operational in January 2019 (x-axis) and monitors that became operational between February 2019 and December 2020 (y-axis). Only monitors with less than 25% missing daily observations in 2021 are included. The correlation coefficients are: (a) 0.847 (b) 0.223.

Figure 2: Monitor clustering within the San Francisco Bay Area




Notes: The green/blue dots show the locations of outdoor/indoor monitors in December 2021. The orange dots represent the locations of official PM$_{2.5}$ monitors. The legend for Panel (b) represents quintiles of California's census tract-level median annual household income in 2019; for Panel (c) it is the average PM$_{2.5}$ concentration in µg/m$^3$ during the period 2019-2021; and Panel (d) represents quintiles of California's census tract-level shares of the Black/Hispanic population in 2019.

Figure 3: Explained variation of added monitors




Notes: Overall $R^2$ (left y-axis) values from census tract-level fixed-effects regressions of the nth (x-axis) monitor added to a tract on all previous monitors added to a tract, separately for the full sample, the above $75^{th}$ percentile sample, and the above $90^{th}$ percentile sample (calculated from the average reading across all 10 monitors). The right y-axis shows the difference in the $R^2$ to the $n^{th}$-1 regression. The data includes daily readings from 2021 from monitors added before 2021. Only monitors with less than 25% missing daily observations in 2021 are included. All regressions include month and day-of-week fixed effects.

# Appendix Tables & Figures

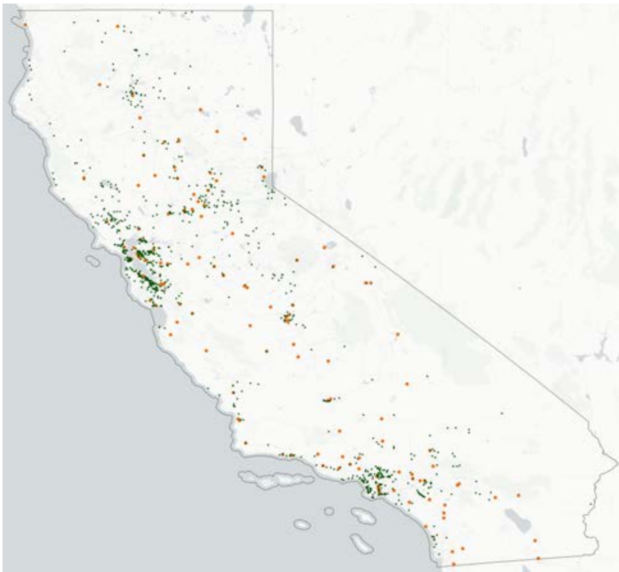
Table A1: Summary statistics

| Variable | All Tracts | (2) Tracts w/ PA monitor | (3) Tracts w/o PA monitor | (2)-(3) |
|---|---|---|---|---|
| Nb. outdoor monitors | 0.940 | 2.517 | 0.000 | 2.517 |
| | (2.351) | (3.290) | (0.000) | [0.047] |
| Nb. indoor monitors | 0.486 | 1.301 | 0.000 | 1.301 |
| | (1.550) | (2.318) | (0.000) | [0.033] |
| Med. annual houshold income, USD | 81,528 | 100,053 | 70,483 | 29,570 |
| | (39,387) | (45,394) | (30,354) | [850] |
| Share Bachelor's degree | 0.333 | 0.447 | 0.265 | 0.182 |
| | (0.212) | (0.212) | (0.180) | [0.004] |
| Share Black population | 0.058 | 0.043 | 0.066 | -0.023 |
| | (0.087) | (0.067) | (0.096) | [0.002] |
| Share Hispanic population | 0.381 | 0.236 | 0.467 | -0.232 |
| | (0.264) | (0.193) | (0.264) | [0.006] |
| Av. $PM_{2.5}$ concentration, µg/m$^3$ | 10.615 | 9.600 | 11.219 | -1.619 |
| | (2.323) | (2.136) | (2.218) | [0.051] |
| Distance to closest EPA monitor, miles | 7.659 | 8.277 | 7.291 | 0.986 |
| | (7.336) | (8.893) | (6.198) | [0.170] |
| Share under 5 y/o | 0.061 | 0.055 | 0.064 | -0.010 |
| | (0.026) | (0.024) | (0.026) | [0.001] |
| Share over 64 y/o | 0.148 | 0.173 | 0.133 | 0.040 |
| | (0.081) | (0.085) | (0.074) | [0.002] |
| Share male | 0.496 | 0.496 | 0.496 | -0.000 |
| | (0.040) | (0.036) | (0.042) | [0.001] |
| Unemployment rate | 0.062 | 0.052 | 0.068 | -0.015 |
| | (0.037) | (0.034) | (0.038) | [0.001] |
| Nb. households | 1,638 | 1,808 | 1,536 | 272 |
| | (732) | (799) | (668) | [17] |
| Share owner-occupied housing units | 0.548 | 0.607 | 0.514 | 0.093 |
| | (0.240) | (0.225) | (0.242) | [0.005] |
| Share single-unit structures | 0.666 | 0.699 | 0.647 | 0.052 |
| | (0.264) | (0.256) | (0.267) | [0.006] |
| Tract size, sq. miles | 19.452 | 37.824 | 8.498 | 29.326 |
| | (161.991) | (244.234) | (77.520) | [3.738] |
| Census tracts | 7,965 | 2,975 | 4,990 | |

Notes: Means and standard deviations (in parentheses) of our dependent variables and all variables included in $X_c$, $P_c$, $S_c$, and $\tilde{X}_c$ at the census tract level. Column 1 includes all tracts, Column 2 includes tracts with at least one monitor present (outdoor or indoor or both), and Column 3 excludes those tracts. Column 4 shows the difference between Columns 2 and 3 and the standard errors (in brackets) from a t-test for the difference in means. The number of monitors refers to the number of operational monitors in December 2021. The average $PM_{2.5}$ concentration covers the years 2019-2021. All demographic variables are from the 2019 American Community Survey.

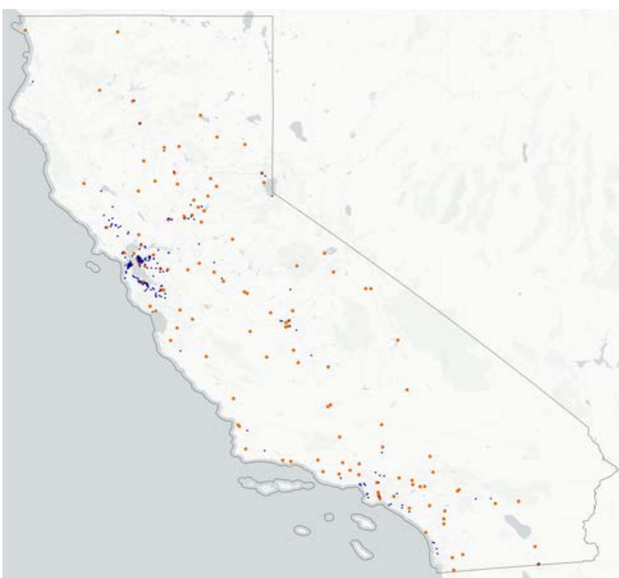
Figure A1: Diffusion of monitors in California

(a) Outdoor monitors | January 2019

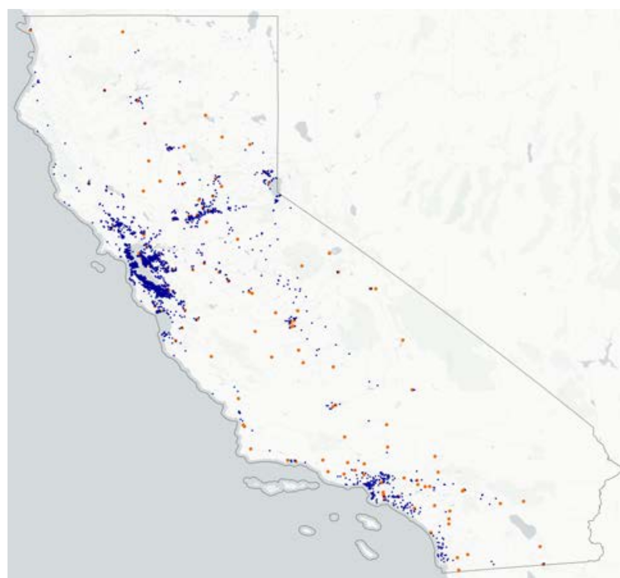


(b) Outdoor monitors | December 2021

(c) Indoor monitors | January 2019



(d) Indoor monitors | December 2021



Notes: The green/blue dots show the locations of outdoor/indoor monitors that were operational in January 2019 (Panels (a) and (c)) and December 2021 (Panels (b) and (d)). The orange dots represent the locations of official $PM_{2.5}$ monitors.