Using Satellite Data to Fill the Gaps in the US Air Pollution Monitoring Network[*]

Daniel M. Sullivan Alan Krupnick

First draft: September 2018
This draft: January 2019

Abstract

Compliance with the Clean Air Act's National Ambient Air Quality Standards (NAAQS) is determined using local pollution monitors. However, most counties have zero or one monitor, and monitors may not represent exposure across wide areas. We use satellite-derived data on fine particulate matter ($PM_{2.5}$) to revisit the compliance determinations for the $PM_{2.5}$ annual NAAQS made in 2015. Mirroring current regulatory practice, we flag counties as "nonattainment" if they contain areas that exceed the NAAQS. Comparing the satellite-based list of nonattainment areas to the official determinations, we estimate that 24.4 million people live in attainment areas that the satellite data suggest should be nonattainment. We then estimate how air quality changes in areas targeted by regulators for improvement, as occurs around ground-based monitors that are designated nonattainment. The estimates suggest that proper classification would have prevented 5,652 premature deaths between 2016 and 2017, a welfare gain of $51 billion.

JEL Codes: Q51, Q53, Q58
Keywords: Environmental Regulation; Clean Air Act; Air Pollution

[*]Resources for the Future, 1616 P St NW, Washington, DC 20036. Email: sullivan@rff.org, krupnick@rff.org. We thank Alex de Sherbinin and David Diner for advice on using satellite data on particulate matter. We are particularly grateful to Aaron van Donkelaar for sharing data. We also thank Mark Nepf for excellent research assistance. We benefited greatly from feedback from Corbett Grainger, Dave Cohen, and seminar participants at WCERE, HAQAST4, and EPA. This research was supported through NASA cooperative agreement number NNX17AD26A with RFF to estimate the value of information obtained from satellite-based remote sensing.

1 Introduction

The Clean Air Act (CAA) is the foundation of air quality regulation in the United States. Under the CAA, the US Environmental Protection Agency (EPA) establishes National Ambient Air Quality Standards (NAAQS) for several air pollutants: $PM_{2.5}$, PM_{10}, ozone, NO_x, SO_2, CO, and lead. The NAAQS themselves are statistics describing concentration levels over various time horizons, e.g., the three-year daily average or the 98th percentile of daily maxima over three years. Air pollution monitors across the country regularly measure concentrations of the regulated pollutants and regulators calculate "design values" from these measurements to compare against the NAAQS. If a monitor reports a design value above the NAAQS, the monitor's jurisdiction (usually the county) is classified as "nonattainment" for that pollutant standard and is subject to a series of requirements to bring its design value within the standard.[1] Past research has found that air quality improves significantly faster in nonattainment areas than in attainment areas, resulting in concomitant health and welfare benefits (see, e.g., Currie et al. 2014; Bishop, Ketcham, and Kuminoff 2018).

But these air quality standards are only as effective as the EPA's monitoring network, which is limited. The majority of US counties lack monitors altogether, and readings at an air pollution monitor do not necessarily represent concentrations across a wide area like a county. Of 3,100 counties in the United States, only 651 (21 percent) have any $PM_{2.5}$ monitors. Of these, 48 percent have a single monitor. In such cases, standard practice is to assume that the concentrations registered by that monitor are representative of concentrations throughout the county. Good monitor placement is obviously critical to this assumption, but recent research shows that some monitors appear to be placed in areas of low pollution relative to elsewhere in the county, such as upwind of major point sources (Grainger, Schreiber, and Chang 2018).

The purpose of this research is to measure how many people live in gaps in the monitoring network where pollution levels are high but undetected by monitors and therefore undetected by regulators. We use high-resolution satellite-derived data ($\sim$1 km^2) on ground-level $PM_{2.5}$ concentrations to find counties that are designated as attainment but contain areas that violate the NAAQS according to the satellite data.

We find that 54 counties in 11 states, home to 24.4 million people, are misclassified

1. There are some additional factors that EPA may consider when making nonattainment designations, which we describe in more detail in Section 2.1.

according to satellite data from the time period EPA used to originally assign attainment designations (2011–2013). Of these, 10.9 million live in counties that do not contain any $PM_{2.5}$ monitors. Reclassifying all these misclassified individuals as nonattainment would more than double the total nonattainment population, currently 23.2 million. We also find that the misclassification rate varies considerably across demographic groups, with rural individuals and black individuals the most likely to be misclassified.

One consequence for these misclassified counties is that their residents did not enjoy the same accelerated improvements in air quality that properly classified nonattainment counties did after their nonattainment designation. We estimate the value of these forgone health benefits by measuring the effect of a new nonattainment designation on pollution concentrations near a monitor. Auffhammer, Bento, and Lowe (2009) find that when regulators take action to bring their counties into attainment, they specifically target high-pollution monitors—the immediate cause of the nonattainment designation—for improvements. We measure the effect of this targeting after the 2015 nonattainment designation using a difference-in-differences design with monitor data and find that $PM_{2.5}$ around targeted monitors (monitors over the NAAQS in nonattainment counties) decreased by 2.3 $\mu g/m^3$ more than it did around attainment monitors, which we take to be the counterfactual of targeted monitors absent the regulatory intervention. Untargeted monitors in nonattainment areas (monitors that did not exceed the NAAQS) also saw relative decreases, though smaller than those of targeted monitors.

To calculate the social benefits of using the satellite data in the regulatory process, we consider the hypothetical where local regulators target *all* areas that exceed the NAAQS instead of just those areas exceeding the NAAQS according to ground-based monitors. Using a standard concentration-response estimate from Lepeule et al. (2012) and the estimated effect of regulators' targeting of monitors, we calculate the excess mortality due to misclassification. We estimate the decrease in $PM_{2.5}$ that misclassified counties would have experienced had they been properly classified and use our estimates to calculate the number of premature deaths that would have been avoided. We find that improved air quality would have prevented 5,652 premature deaths in misclassified areas between 2016 and 2017. Using the value of a statistical life (VSL), this would result in $51 billion in benefits to misclassified counties had states acted as quickly to reduce $PM_{2.5}$ levels in these areas as they have in nonattainment areas.

2 Air Quality Regulation with Sparse Monitoring

2.1 Background on the Clean Air Act

The Clean Air Act (CAA) of 1970 and its subsequent amendments form the basis of current air quality regulation in the United States (Revesz 2015).[2]

The CAA directs the administrator of the EPA to issue NAAQS for certain pollutants. Limits must be imposed on air pollutants that "may reasonably be anticipated to endanger public health or welfare" (42 USC § 7408(a)(1)(A)), and these limits should be set to "protect the public health" with "an adequate margin of safety" (§ 7409(b)(1)).[3]

The NAAQS are supposed to be reviewed and possibly revised no later than every five years, although this schedule is rarely met. These assessments are made in consultation with an independent scientific review committee, the Clean Air Scientific Advisory Committee (CASAC), which provides scientific and technical advice.

The current primary standards for $PM_{2.5}$ were set in 2012 as (1) an annual average of no more than 12 micrograms per cubic meter ($\mu g/m^3$) (down from 15 $\mu g/m^3$ in 2005) and (2) a 98th percentile of daily readings no more than 35 $\mu g/m^3$ (unchanged from 2005). Both metrics are calculated using the three most recent years of monitor data. As we discuss further in Section 5, this study focuses on the annual standard of 12 $\mu g/m^3$.

Once a NAAQS is established for a pollutant, each state formally recommends to EPA which areas (generally counties) should be classified as in attainment with the NAAQS, which should be classified as nonattainment areas (§ 107(d)(1)), and which are nonclassifiable (effectively attainment) areas. The state is required to use the latest three years of monitoring data to do this, but it also may use atmospheric modeling, emissions inventories, and other tools.[4] States also identify areas that contribute to

2. In general, see Revesz (2015), chapter 5, for detailed history and review of air quality regulation in the United States at the federal level.

3. § 7408(a)(1)(B) also requires that the presence of the pollutant be due to "numerous or diverse mobile or stationary sources." The public health standard prescribed in § 7409(b)(1) is known as the "primary" standard. § 7409(b)(2) provides for "secondary" standards that "protect the public welfare." Welfare "includes, but is not limited to, effects on soils, water, crops, vegetation, manmade materials, animals, wildlife, weather, visibility, and climate, damage to and deterioration of property, and hazards to transportation, as well as effects on economic values and on personal comfort and well-being" (§ 7602(h)). Note, however, that the costs of meeting the standards may not be considered when setting standards.

4. When the $PM_{2.5}$ standards were set in 2012, EPA directed states to use air quality monitoring data from 2010 through 2012 in their initial recommendations for nonattainment areas and said it would use data from 2011 through 2013 in its final determination if 2013 data were available in time.

downwind air quality violations and include them in their nonattainment recommendations. EPA examines the state submission and then makes its determination on nonattainment designations, which states can appeal.

The CAA also permits reclassifications of an area's status as conditions change. Most reclassifications are made after states or other groups petition to move from nonattainment status to attainment. EPA can also reclassify in the other direction on its own or if asked to by petitioners, but this rarely happens. EPA data show that only one area has ever been reclassified from attainment status to nonattainment for any $PM_{2.5}$ standard.[5]

Once an area is officially designated nonattainment, the state or states in which the area is located must submit a state implementation plan (SIP) to EPA that outlines how the NAAQS will be met (e.g., what restrictions will be placed on which industries in which parts of the state). Polluters in nonattainment areas face more stringent regulations than those in attainment, such as a requirement to use the best available control technology. Areas with more severe nonattainment designations face tighter restrictions but have longer deadlines to reach attainment. States that continually fail to make "reasonable further progress" in reaching attainment may face federal funding sanctions or other consequences. Understanding the schedule and speed of nonattainment areas in reaching attainment is important when we estimate the health benefits of making proper designations.

2.2 Problems with a Limited Network of Air Pollution Monitors

The NAAQS and the attainment designations depend on EPA's pollution monitors to provide an accurate measure of how much pollution people are exposed to. However, there are several problems with using a limited network of monitors to measure exposure to a spatially dispersed population. The fundamental issues are that (1) air pollution varies significantly over short distances, with spikes around every factory, every road, and every refinery;[6] and (2) pollution can travel long distances from its source. As such, the correlation between a monitor's readings and concentration on the ground decays quickly with distance; the farther you get from a monitor, the less the

5. Pinal County, Arizona, was reclassified in 2011 as nonattainment with the $PM_{2.5}$ 2006 rule (24-hour standard) two years after initial classifications were made for that rule in 2009. Reclassification from attainment to nonattainment for any pollutant is rare. It has happened only 59 times across all 13 criteria pollutant standards. Of these, 34 percent were changes to counties' SO_2 status (2010 rule) and occurred in the last three years.

6. See, e.g., Hu et al. (2009).

monitor can tell you. But in 2015, 79 percent of counties did not have a $PM_{2.5}$ monitor, 10 percent had one monitor, 5 percent had two monitors, and 6 percent had three or more.

In addition to physical problems of measuring pollution, there are two problems with regulatory air pollution monitors that arise because air pollution is generated by economic activity. The first problem is Goodhart's law: once a metric of economic activity becomes a regulatory target, it is no longer a good metric of the underlying activity. This is because economic actors may adapt their behavior to affect the metric, like when a teacher changes his curriculum to better fit the standardized tests used to evaluate his performance. For air pollution, this "teaching to the test" could happen in a number of ways: strategically placing monitors in less-polluted parts of the county, strategic timing of abatement by polluters when monitors are in operation, or the gradual relocation of polluters over time from locations upwind of monitors to locations downwind. Grainger, Schreiber, and Chang (2018) present evidence of strategic monitor placement. Zou (2018) presents evidence that firms reduce their pollution on days that $PM_{2.5}$ and PM_{10} monitors are in operation, since some monitors operate only on select days and their schedule is published in advance.[7] Our paper does not test for these problems directly, but our results are likely driven in part by them.

The second problem is that monitors are fixed in space and time, while the location of pollution is constantly changing. A pollution monitor provides a sample concentration from a single point in what could be a large area with varying topography, wind conditions, traffic patterns, and density of industry. Since air quality regulation is ultimately aimed at improving health, how people, pollution, and polluters are differentially distributed across space cannot be ignored. Furthermore, none of these distributions are static. As polluters in different locations change their polluting behavior and as establishments relocate or open for the first time, the spatial distribution of emissions changes, and in turn the overall exposure to nearby residents changes. As neighborhoods grow in some parts of a city and shrink in others, overall exposure changes. As the local climate becomes hotter, airborne pollutants react in different ways and total exposure changes. Meanwhile, the monitor observes the air at

7. Intermittent operation is primarily a problem with older monitors and PM_{10} monitors. In our sample, 56 percent of $PM_{2.5}$ monitors gathered data on fewer than 121 days in 2015, and 23 percent of those gathered data on fewer than 80 days.

the same physical location and is blind to all these changes. If a monitor initially corresponds to the median concentration in a county, a year later it may correspond to the 40th percentile, the 60th percentile, or some other order statistic from the population exposure distribution.

These problems together motivate a closer look at how well the monitors in the United States measure resident exposure to air pollution and, in turn, at whether the process of designating areas can be improved. A natural next question is ask is whether the county is the best unit of designation. Given how much pollution can vary over short distances, it may be more efficient to determine attainment status for areas smaller than the county. This would reduce the number of people living under a nonattainment designation who themselves have air quality within the NAAQS. However, abatement becomes harder at smaller scales—imagine a single residential neighborhood or apartment building being designated nonattainment and tasked with reducing their local pollution. This problem of the optimal unit of designation is separate from whether satellite data can improve the existing regulatory system. As such, we leave it for a future paper.

3 Research Design

Are there areas exceeding the NAAQS that EPA's monitor network has missed? What might the mortality benefits have been had these areas been correctly classified as nonattainment?

To answer the first question, we compare the satellite-derived data on ground-level $PM_{2.5}$ with official nonattainment designations to flag census blocks which (1) have satellite $PM_{2.5}$ readings over the NAAQS and (2) are classified as attainment/unclassifiable. We refer to counties with any such areas as "misclassified," since they would have been classified as nonattainment had EPA's monitor network had the same spatial coverage as the satellite data, i.e., one monitor for every square kilometer.

Our primary answer to the second question takes a two-step approach. First, we estimate the effect that regulators have on local air quality when they target an area for improvements. A nonattainment designation pushes local regulators to lower pollution in order to get their county to attainment status. Past research has found that regulators can be effective at targeting monitors that cause their county to be in nonattainment. As Auffhammer, Bento, and Lowe (2009) observe, regulators in nonattainment areas have less incentive to target monitors that are not over the NAAQS because the nonattainment designation depends on readings at the highest monitor.[8] We take an empirical approach similar to that of Auffhammer, Bento, and Lowe (2009); Grainger (2012); and Bento, Freedman, and Lang (2015) to estimate the effect of $PM_{2.5}$ nonattainment status on violating monitors. We estimate a difference-in-differences regression to measure the effect on $PM_{2.5}$ concentrations over time for monitors in nonattainment areas that register readings over the NAAQS (termed Group I) versus monitors in nonattainment areas that register concentrations below the NAAQS (termed Group II) versus monitor readings in attainment areas (termed Group III):

$$\begin{aligned} P_{mt} = {} & \beta_1 \left(\text{Nonattainment}_m \times \text{Over NAAQS}_m \times \text{post}_t\right) + \\ & \beta_2 \left(\text{Nonattainment}_m \times \text{Under NAAQS}_m \times \text{post}_t\right) + \\ & \delta_t + \delta_m + \varepsilon_{mt} \end{aligned} \tag{1}$$

where P_{mt} is the pollution reading for monitor m at time t. The indicator variables δ_t and δ_m control for year and monitor effects. Nonattainment and "over NAAQS" status are taken from the year 2015, the first year nonattainment determinations were made for the 2012 $PM_{2.5}$ rule. Here β_1 captures the change in targeted monitors in nonattainment areas (Group I) relative to changes in attainment monitors (Group III). Likewise β_2 captures the change in untargeted monitors in nonattainment areas (Group II) relative to Group III. Because these monitors are not targeted, we would anticipate β_2 to have a smaller magnitude than β_1, with much of the effect on untargeted monitors coming as a side effect from efforts to reduce pollution at the nearby targeted monitors. (Because a counties in our sample were not designated nonattainment without a violating monitor, all Group II monitors must have a Group I monitor in the same county.)

We invoke the usual parallel trends assumption that, absent the regulatory intervention, the trends of Group I and Group II monitors would have continued to follow the trend of Group III monitors. Under this assumption, $\hat{\beta}_1$ and $\hat{\beta}_2$ measure the casual effect of regulators' response to a nonattainment designation.

8. "The federal regulation creates an incentive for the local regulator to closely track the monitors that put the county at 'risk' of becoming out of attainment. The regulator then allocates effort in terms of monitoring and enforcement activities to the different monitors by comparing the future costs of getting out of attainment to the present costs associated with the reduction in the emissions around 'risky' monitors. The resulting equilibrium is a schedule of heterogeneous monitoring and enforcement efforts such that more effort is allocated to dirtier monitors" (Auffhammer, Bento, and Lowe 2009, p. 17).

After estimating $\hat{\beta}_1$ and $\hat{\beta}_2$, we consider a scenario where federal regulators have monitors for every square kilometer. This gives local regulators the incentive to target all areas that exceed the NAAQS instead of one or two specific points as in the current case with sparse monitoring. We assume that if areas over the NAAQS in misclassified counties had been correctly classified, they would have experienced declines in $PM_{2.5}$ similar to those experienced by targeted monitors after their nonattainment designation, i.e., $\hat{\beta}_1$. Similarly, we assume areas under the NAAQS in misclassified counties would get the spillover improvement $\hat{\beta}_2$. We then calculate excess mortality using the estimate of concentration-mortality response from Lepeule et al. (2012) which is commonly used in EPA Regulatory Impact Analyses (see, e.g., EPA 2012; 2014; 2015). We define excess mortality as the number of deaths would have been avoided if misclassified counties had been classified as nonattainment.

We also go through a secondary, much simpler exercise to calculate excess mortality that ignores regulator behavior and abatement spillovers. In the non-behavioral method, we consider only the health benefits of bringing misclassified areas exactly in line with the NAAQS. Specifically, we estimate what health benefits would have been if areas over the NAAQS in misclassified areas had their $PM_{2.5}$ decreased exactly to the NAAQS (12 $\mu g/m^3$) while all areas under the NAAQS—including those in misclassified areas—follow their current trends. After calculating the potential air quality improvements, we again use the concentration-response from Lepeule et al. (2012) to calculate excess mortality and its social cost.

We focus on the most recent revision to the $PM_{2.5}$ NAAQS, which was made in 2012, lowering the limit for annual average $PM_{2.5}$ to 12 $\mu g/m^3$. States submitted recommendations for their nonattainment designations in 2014 using monitor data from 2011 to 2013, and official designations were announced in 2015. We focus our regression analyses on the five year window around 2015, from 2013 through 2017. We do this for a few reasons. First, we want to use a balanced panel of monitors in the regression so that monitors leaving or entering the sample do not bias the results. After restricting to monitors in continuous operation from 2013 through 2017, we have 14 monitors in Group I, 49 in Group II, and 852 in Group III.[9] Increasing the time span of the sample increases the length of the pre-treatment period, but also removes monitors

9. Here we really mean "continuously according to schedule," so a monitor that was supposed to operate every sixth day is considered to have "continuously" operated if it did so for the entire period in question.

from the sample. The second reason for this restriction is that a new daily standard for $PM_{2.5}$ was implemented in 2009, and we want to limit the influence of this event on our estimates.

4 Data

4.1 Air pollution monitors and attainment designations

Data on EPA's air pollution monitoring system come from EPA and cover every air pollution monitor from 1999 through 2017. The data include latitude and longitude, days of operation, pollution readings, and whether the monitor can be used to determine NAAQS compliance. When used, average annual readings for each monitor exclude concurred exceptional events.[10]

Table 1 reports the number of monitors available for NAAQS compliance during our period of study. Panel A reports the number of monitors designated as NAAQS primary compliance monitors that operated in the given year. Column 1 reports how many monitors operated no more than 80 days during that year, column 2 reports how many operated 81–120 days, and so on. The strongest time trend is the addition of monitors operating more than 300 days a year. Panel B reports how many monitors had sufficient data over the prior three years to calculate a design value. For any given year and frequency, the number of monitors with three years of data is generally less than those with one year of data, though small anomalies can occur when monitors move across the frequency categories year to year. Both panels show that a significant proportion of monitors operate less than once every three days (columns 1 and 2). Even as late as 2017, 10 percent of monitors available to calculate design values to compare with the NAAQS operated no more than 80 days per year.

Data on attainment status are taken from EPA's Green Book.[11]

10. In exceptional events that are outside state regulators' control, such as wildfires, state regulators may petition to have monitor readings from those events excluded from design value averages. EPA then chooses whether to concur that the event was exceptional and allow it to be excluded from the design value.

11. See https://www.epa.gov/green-book/green-book-data-download.

4.2 Satellite-derived concentration data

The satellite-derived $PM_{2.5}$ concentration data come from a variety of sources (van Donkelaar et al. 2015; van Donkelaar et al. 2016).[12] The data are primarily gathered by satellite-based instruments that measure aerosol optical depth (AOD). The best known of these instruments among economists are the MODIS instruments aboard the Terra and Aqua satellites (see, e.g., Zou 2018; Grainger, Schreiber, and Chang 2018; Gendron-Carrier et al. 2018). As these satellites orbit Earth, the MODIS instruments on board capture data on the density of airborne particles. It does this by comparing the intensity of solar radiation at the top of the atmosphere with how much radiation is reflected by Earth's surface. The more airborne particles there are to scatter and absorb this radiation, the less radiation is reflected to the satellite.

Both satellites follow a polar orbit, going from the North Pole to the South Pole and back to the North Pole every 100 minutes or so. As the satellites orbit pole to pole, Earth continues to rotate, giving the satellites a new swath of ground to scan. The satellites' orbits are calculated so that they pass over and scan any given point on Earth at approximately the same time every day. On the sun-facing side of Earth, Terra crosses the equator at approximately 10:30 a.m. local time with each orbital pass, while Aqua crosses the equator at approximately 1:30 p.m. Thus, every location is scanned by each satellite approximately once per day at roughly the same time every day. These once-a-day readings are temporally sparser than hourly readings available from ground monitors. However, as discussed in the previous section, few monitors report hourly data, and most do not collect data every day.

Van Donkelaar et al. (2015) and van Donkelaar et al. (2016) combine the AOD data from the MODIS MISR (also aboard Terra) and SeaWIFS (aboard OrbView-2 satellite) instruments with results from the chemical transport model GEOS-Chem. GEOS-Chem provides information about how pollutants are transported from one area to another by the wind and how chemical compounds change as they travel. This combination of measurements and simulation is calibrated using ground-based monitored observations of $PM_{2.5}$ at a monthly timescale. The data are then averaged by year for every 0.01-by-0.01-degree grid cell, which is approximately 1 km^2 in area.

While satellite-derived data on air pollution provide unique opportunities for

12. The data we use here are an updated version of the North America data developed in van Donkelaar et al. (2015), which uses advances presented in van Donkelaar et al. (2016). We are very grateful to Aaron van Donkelaar for giving us access to these data.

researchers and policy makers, they also come with a few caveats. First, the satellites do not measure $PM_{2.5}$ directly. They measure AOD, which must be scaled to $PM_{2.5}$ based on local conditions. This is not altogether straightforward even for researchers with atmospheric sciences training, but AOD itself can sometimes be used to estimate comparative $PM_{2.5}$ levels in localized areas. However, when comparing data with general policy thresholds such as the NAAQS, they must be accurately scaled. Second, satellites cannot measure ground-level conditions on cloudy days. This is one reason the satellite-derived data are more reliable at large timescales (months or years) than small ones (hours or days). Third, the accuracy of the data depends on the sample of ground-based monitors used for calibration. For example, data calibrated globally could have mean-zero error globally, but sub-samples of the data (e.g., the data for North America) may not be mean zero. To avoid this problem, we use data specifically calibrated for North America, which are quite accurate.[13]

Figure 1 plots the correlation between the van Donkelaar et al. satellite-derived data (vertical axis) and annual average readings from ground-based monitors (horizontal axis). The satellite data for each monitor is taken from the 0.01-by-0.01-degree cell in which the monitor is located. Faint markers indicate individual monitor–grid cell pairs; bold markers indicate the average for every bin centered at integers on the horizontal axis (i.e., satellite average for monitor readings of 1 ± 0.5 $\mu g/m^3$). The shape and color of each marker indicates how frequently the monitor operates: red circles for monitors that operate no more than 80 days per year; yellow triangles for 81–120 days; blue squares for 121–300 days; and green pentagons for those operating more than 300 days per year. Dashed gray lines show the 12 $\mu g/m^3$ NAAQS threshold for nonattainment classification. In general, the satellites show strong agreement with the monitors, especially at lower monitor readings. At higher monitor readings, the satellites tend to underestimate pollution concentrations relative to the monitors. This would imply that our methodology may be somewhat conservative in determining areas that are misclassified as attainment.

4.3 Population and demographic data

Block-level data on population counts and race/ethnicity come from the 2010 census. Block group–level data on educational attainment and household income come from

13. Compare correlation between monitors and satellite data calibrated for North America shown in Figure 1, discussed below, and the equivalent figure for globally calibrated data restricted to North America in Figure A1, which shows a systematic upward bias relative to the monitors.

the 2005–2010 American Community Survey (ACS). Data on county-level all-cause mortality come from the Centers for Disease Control and Prevention's (CDC's) Compressed Mortality File.[14]

5 Results

5.1 Monitor coverage and nonattainment status

We begin by looking at the locations of $PM_{2.5}$ monitors in the continental United States. Figure 2 shows the location of each of the monitors that were used to make the 2015 attainment determinations. It also labels monitors based on how many days per year the monitor is required to operate following the same scheme as in Figure 1. We see significant heterogeneity across states in both the density of monitors and the frequency of their use. Some states have dense monitor networks that operate daily or near daily (e.g., California, Pennsylvania). Others are hardly monitored at all (e.g., Montana, Maine, Mississippi, Nebraska, Nevada, Idaho).[15] Still others have many monitors, but each of those monitors does not operate more than 80 days per year (e.g., Wisconsin, Wyoming). Even within states, coverage can vary. Most of California is densely monitored, but in central California the monitors operate nearly every day, while in the Los Angeles Basin they operate no more than once every six days.

Figure 3 shows the designated nonattainment areas established in 2015 for the annual $PM_{2.5}$ standard promulgated in 2012. These nonattainment areas cover central California; the Los Angeles Basin; West Silver Valley, Idaho; Cleveland; Pittsburgh; and Philadelphia. All these areas were designated as nonattainment because of high monitor readings and not because they contributed to a downwind nonattainment area.

The key question the satellite data can answer is whether other counties also exceed the NAAQS.

5.2 Satellite-measured concentrations and misclassified areas

Figure 4 shows the difference between the satellite-measured three-year average $PM_{2.5}$ concentrations and the annual NAAQS for the continental United States in 2015. Blue

14. See https://wonder.cdc.gov/cmf-icd10.html.

15. Illinois is a special case where the monitors present in the state were deemed by EPA to be of insufficient quality to be used for NAAQS assessment. Therefore, all the data were thrown out, and the entire state was classified as attainment/unclassifiable.

areas correspond to those for which the satellite predicts a design value below the NAAQS (12 μg/m^3), red areas are those predicted above the NAAQS, and white areas are those right at the NAAQS. This map shows many areas above the NAAQS, particularly in California, where Figure 3 also showed large nonattainment areas. However, Figure 4 also suggests that a large share of the Midwest is close to the NAAQS, and several areas have concentrations well above the NAAQS despite being classified as attainment areas.

Figure 5 focuses on some of these hot spots, highlighting the area bounded by Chicago on the north and west, Louisville on the south, and Pittsburgh on the east. Again, red corresponds to concentrations over the NAAQS, white about equal to the NAAQS, and blue under the NAAQS. County boundaries are plotted in white, official nonattainment areas are bounded in orange, and monitor locations are represented by black dots.

This map gives several examples of misclassified areas. First are areas with concentrations exceeding the NAAQS but that have no monitors in their counties (e.g., Logansport, Indiana, north of Indianapolis and southeast of Chicago). Second are areas exceeding the NAAQS that are not detected because the monitors are too far away from the hot spots (e.g., southeast of Logansport). Third are cities with multiple monitors, but those monitors are located on the edges of the hot spots and miss peak concentrations (e.g., Indianapolis, Louisville, and Cincinnati).

Figure 6 shows all the misclassified counties—those that were designated as attainment but contain areas that exceed the NAAQS—in the continental United States. There are 54 such counties across 11 states.[16] Table 2 lists the number of people living in misclassified counties in each state, with separate counts for counties with monitors and without. For counties that include areas that are both attainment and nonattainment, we treat the attainment part of the county as a distinct county.[17] All told, 24.4 million people live in misclassified areas. Of these, 10.9 million live in counties with no monitors. The states with the largest populations of unmonitored and misclassified people are Illinois (6.4 million misclassified, all unmonitored); California (4.9 million misclassified, of which 0.8 million are unmonitored); and Texas (4.5 million misclassified, with 0.4 million unmonitored). Two other states have sizable

16. Table A1 lists all misclassified counties and their core-based statistical areas.

17. If the satellite data find that the attainment portion should also be nonattainment, we count only people living in the attainment portion as misclassified. Similarly, if the nonattainment portion has pollution monitors but the misclassified attainment portion does not, we describe the attainment "county" as having no monitors.

misclassified populations that are unmonitored: Kentucky (1.2 million misclassified, 1 million unmonitored) and Ohio (2.2 million misclassified, 0.9 million unmonitored). The total number of people living in misclassified counties is 24.4 million, slightly more than the number of people currently living in official nonattainment areas (23.3 million).[18]

5.3 Distribution of people across attainment groups

Table 3 summarizes how various demographic groups are distributed across correctly classified attainment areas and nonattainment areas, as well as misclassified areas. To calculate these figures, we use 2010 census block-level data on population, race/ethnicity, and share urban, and 2005–2010 American Community Survey (ACS) block group–level data on education and income. The first three columns show what percentage of the listed group resides in each type of classification area; these columns sum to 100 by construction. For example, the first row shows that of the 306.6 million people in our sample, 84.5 percent were correctly classified as attainment, 8.0 percent (24.4 million people) were misclassified as attainment, and 7.6 percent (23.2 million) were classified as nonattainment. Column 4 shows the percentage of the given group that live in an area that satellite data show should be nonattainment (the sum of columns 2 and 3). Column 5 shows the ratio of people misclassified by monitors to the number of people flagged by satellites as nonattainment (column 2 divided by column 4). We refer to this ratio as the false negative rate of the monitoring network. For the whole population, the satellite data found that 15.5 percent (column 2 plus column 3) should have been classified as nonattainment (misclassified population plus officially nonattainment), but of these over half (8 percent over 15.6 percent) were misclassified, for a false negative rate of 51 percent.

Similarly, the second row of Table 3 shows that the false negative rate in rural areas is 68.6 percent (2.4 percent divided by 3.6 percent). By contrast, urban areas (third row) have a false negative rate was 50.5 percent. This false negative rate is likely lower than that for rural areas because monitors are more concentrated in urban areas and are thus more likely to detect when urban areas exceed the NAAQS compared to rural areas. Similarly, pollution concentrations are likely to be higher in urban areas where pollution is more concentrated.

18. We include in these counts Chicago (and the rest of Illinois) and Houston whose designation is attainment/unclassifiable because their monitoring was not deemed reliable enough to determine NAAQS compliance.

There are also large disparities in both pollution concentrations and monitors' false negative rates across racial and ethnic groups. White (non-Hispanic) individuals are the least likely to live in areas that the satellites flagged as nonattainment (misclassified or official nonattainment), with 11.3 percent living in such areas. The monitors' false negative rate for whites is 59.6 percent. The fraction of black individuals flagged by the satellite as nonattainment is 14.6 percent, with a false negative rate of 65.1 percent. Hispanics as a group are the most likely to live in an area that exceeds the NAAQS, with 30.0 percent living in these areas. However, their false negative rate is somewhat lower than that of whites and blacks at 38.5 percent. Asians have a comparable rate of living in nonattainment and misclassified areas (26.3 percent) but have the lowest false negative rate of all groups (32.8 percent).

There are similar disparities across educational attainment groups, which are defined as the highest level of education attained by people at least 25 years old. Those without a high school diploma are the most likely to live in areas exceeding the NAAQS (19.0 percent), but the least likely to be missed by monitors (45.0 percent false negative rate). Those with only a high school diploma are less likely to live in an exceedance area (13.3 percent) but more likely to be missed by monitors (54.5 percent). Those with a partial college education are similar (14.9 percent over the NAAQS, 51.4 percent false negative rate) as are those with a college degree or more (15.3 percent over the NAAQS; 53.3 percent false negative rate).

Looking at households by income, there is surprisingly little variation in both their likelihood of living in an exceedance area and their false negative rate.

5.4 Excess deaths from being misclassified as attainment

A nonattainment designation requires local regulators to develop and implement plans to bring their jurisdiction into attainment.[19] Had misclassified counties been designated nonattainment, they might have experienced health benefits from the ensuing improvement in air quality. In this section, we estimate the extent of these unrealized health benefits in two ways, first considering how regulator behavior affects air quality, and then without such considerations.

19. See Chay and Greenstone (2003), Currie et al. (2014), and Bishop, Ketcham, and Kuminoff (2018).

5.4.1 Effect on Air Quality from Regulator Behavior

As discussed in Section 3, the structure of the CAA and NAAQS create an incentive for regulators to specifically target improvements around those monitors that exceed the NAAQS rather than across their jurisdiction as a whole (Auffhammer, Bento, and Lowe 2009; Grainger 2012; Bento, Freedman, and Lang 2015). Following past research, Equation (1) outlines our difference-in-differences strategy to estimate the effect of this behavior on air quality.

Before estimating Equation (1), we want to assess the validity of the parallel trends assumption by estimating the event study equivalent of Equation (1),

$$\begin{aligned} P_{mt} = &\sum_{y \neq 2015} \beta_{1,y} \left(\text{Nonattainment}_m \times \text{Over NAAQS}_m \times \mathbf{1}\{t = y\}\right) + \\ &\sum_{y \neq 2015} \beta_{2,y} \left(\text{Nonattainment}_m \times \text{Under NAAQS}_m \times \mathbf{1}\{t = y\}\right) + \\ &\delta_t + \delta_m + \varepsilon_{mt} \end{aligned} \tag{2}$$

which takes attainment monitors (Group III) as the baseline and estimates how the trends of Group I and Group II deviate from Group III. For example, the coefficient $\beta_{1,2016}$ is the difference between Group I in 2016 and 2015, minus the difference in Group III between 2016 and 2015. If $\beta_{1,2016} = 0$, then Group I and Group III followed the same trend between 2015 and 2016. If $\beta_{1,2016} < 0$, Group I saw larger improvements in air quality. The coefficients for 2015 are omitted, so each coefficient is the change relative to 2015.

Figure 7 plots the $\hat{\beta}_{1,y}$ and $\hat{\beta}_{2,y}$ estimates and their 95 percent confidence intervals. The Group I coefficients (red solid line) are not statistically significant before 2015, showing a slight downward trend relative to the Group III monitors. The large error bands are likely due to the small number of monitors in Group I (14 monitors) and the different local shocks faced by each monitor. However, in 2016, the first year after the new nonattainment desigations, the confidence interval narrows considerably and the trend takes a slightly negative turn. The 2016 coefficient is 66 percent larger in magnitude than the 2014 coefficient, but this difference is not statistically significant. The Group I monitors make a considerable break from trend in 2017, with a decrease of 3.3 $\mu g/m^3$ $PM_{2.5}$ relative to Group III. The Group II coefficients are also somewhat noisy, hovering around zero with a slight drop in 2017.

Table 4 reports the difference-in-difference regression results. Column 1 reports the pooled diff-in-diff of monitor readings on year indicators and a single nonattainment–post interaction. The “post” period is 2016 and 2017, following the nonattainment designations made in 2015. The preferred specification in column 2 adds a variable identifying monitors in nonattainment areas operating in 2016–2017 and showing violation of the NAAQS (nonattainment–post–over NAAQS), a triple interaction term. Both regressions also account for any idiosyncrasies attributable to specific monitors through monitor-level fixed effects. All standard errors are clustered by monitor.

Column 1 shows that the average monitor in nonattainment areas records 1.02 $\mu g/m^3$ less $PM_{2.5}$ after the 2015 nonattainment designations relative to monitors in attainment areas. Column 2 allows for a separate effect on monitors over the NAAQS and shows that the overall decrease in pollution is being driven mostly by monitors over the NAAQS. Under-NAAQS nonattainment monitors see pollution drop 0.64 $\mu g/m^3$ relative to attainment monitors, while over-NAAQS monitors see a 2.35 $\mu g/m^3$ decrease.

5.4.2 Calculating Excess Mortality

To calculate excess mortality from misclassification, we suppose that if misclassified areas had been designated as nonattainment, they would have experienced an average decline in pollution levels similar to that in areas properly designated as nonattainment. Specifically, areas over the NAAQS would have their $PM_{2.5}$ decrease by an additional 2.35 $\mu g/m^3$, while $PM_{2.5}$ in non-exceeding areas in the same county would decrease by 0.64 $\mu g/m^3$. We translate these pollution decreases into decreased mortality risk by multiplying them by the concentration-response coefficient from Lepeule et al. (2012) of 14 percent increase in all-cause mortality per additional 10 $\mu g/m^3$ $PM_{2.5}$. This leads to a 3.3 percent increase in all-cause mortality in areas over the NAAQS and a 0.8 percent increase in mortality in the rest of the county. Finally, we multiply these figures by county-specific death rates from the CDC and block-level population from the census.

We find that misclassified counties would have avoided 2,826 premature deaths per year had they been correctly classified. Using EPA’s standard VSL of $9 million, the social cost of this excess mortality is approximately $25.4 billion per year (EPA 2016, p. 4-16). While the excess mortality effect is measured annually, it eventually gets eliminated as pollution trends in nonattainment areas equalize with those in attainment areas. A conservative assumption would be that the trends shown in Figure 7 will equalize after 2017 (after our regression sample) so that all benefits are realized in 2016 and 2017. This would imply that total excess mortality of misclassification was 5,652, with a social cost of $51 billion.

We can also make back-of-the-envelope calculations of excess mortality that ignore regulator behavior. One way is to scale concentrations in misclassified counties so that all areas are under the NAAQS (12 $\mu g/m^3$) and then calculate the mortality savings.[20] This results in an average decrease in $PM_{2.5}$ of 0.5 $\mu g/m^3$ across all misclassified blocks, leading to 1,398 fewer deaths per year, a welfare gain of $12.6 billion per year. A second way is "peak shaving", the scenario where areas that exceed the NAAQS and are located in misclassified counties have their $PM_{2.5}$ decreased to exactly 12 $\mu g/m^3$ while all other areas are left unchanged.[21] This would lead to an average decrease in exposure in over-NAAQS areas (the Group I equivalent) of 0.43 $\mu g/m^3$, and no change in the Group II and Group III equivalents by construction. The resulting excess mortality is 261 deaths per year, with a social value of $2.4 billion per year.

6 Conclusion

The Clean Air Act is the primary air quality regulation in the United States. However, its success in improving health and environmental quality depends on a limited network of stationary pollution monitors to provide regulators with information about local pollution levels. If pollution levels in an area exceed the NAAQS but there is no monitor nearby, that area is unlikely to exercise mitigation actions to reduce its pollution. In this paper, we have used satellite data to provide evidence that significant portions of the country are indeed exceeding the annual $PM_{2.5}$ NAAQS standard but are nevertheless designated as being in attainment. Estimates of how regulators impact air quality in response to a nonattainment designation suggest that correctly classifying misclassified areas could have save thousands of lives, a potential welfare gain to society of $51 billion for 2016 and 2017. A more conservative estimate, ignoring

20. Specifically, let the new hypothetical concentration in block b be $\tilde{x}_{b,c} = x_{b,c}\left(12/x_c^{\max}\right)$ where $x_c^{\max}$ is the maximum concentration in county c.

21. Following the notation in Footnote 20,

$$\tilde{x}_{b,c} = \begin{cases} 12 & \text{if } x_{b,c} > 12 \\ x_{b,c} & \text{else} \end{cases}$$

regulator behavior, values the gains from satellite-corrected classification at $2 billion per year.

While the value to health and social welfare provided by satellite information on air quality appears to be very large, a few caveats are in order. The main caveat is that while satellite data are far more spatially dense than ground-based monitoring data, the conversion of what is actually measured by the satellites (aerosol optical depth) to $PM_{2.5}$ is not without error or bias when compared with monitor readings at the same place. In our case, the bias works to make our surprisingly large estimates of the misclassified population conservative. The other caveat is the relative temporal sparseness of reliable satellite data. To achieve high spatial resolution (while maintaining accuracy) requires aggregating to larger time scales. Yet, on a more macro scale, the satellites provide data for every day, while in 2016 at most 37 percent of monitors were operating daily.

When the CAA first became law in 1970, legislators could not have envisioned the capability of measuring air quality on a spatially precise basis from satellites. Our results suggest that EPA should examine whether there is scope for satellite data to be used as one of the factors that enter into the designation decision. Alternatively, satellite data can be used as a guide in monitor placement and a check that vulnerable populations are being protected by the Clean Air Act. Failing that, the Clean Air Act should be reopened to change the designation process to better protect the health of the US population.

References

Auffhammer, Maximilian, Antonio M. Bento, and Scott E. Lowe. 2009. "Measuring the effects of the Clean Air Act Amendments on ambient PM10 concentrations: The critical importance of a spatially disaggregated analysis." *Journal of Environmental Economics and Management* 58:15–26.

Bento, Antonio, Matthew Freedman, and Corey Lang. 2015. "Who Benefits from Environmental Regulation? Evidence from the Clean Air Act Amendments." *Review of Economics and Statistics* 97 (3): 610–622.

Bishop, Kelly C., Jonathan D. Ketcham, and Nicolai V. Kuminoff. 2018. "Hazed and Confused: The Effect of Air Pollution on Dementia." NBER Working Paper #24970. http://www.nber.org/papers/w24970.

Chay, Kenneth Y., and Michael Greenstone. 2003. “Air Quality, Infant Mortality, and the Clean Air Act of 1970.” NBER Working Paper #10053, Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w10053.

Currie, Janet, Joshua Graff Zivin, Jamie Mullins, and Matthew Neidell. 2014. “What Do We Know About Short- and Long-Term Effects of Early-Life Exposure to Pollution?” *Annual Review of Resource Economics* 6 (1): 217–247.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A. Turner. 2018. “Subways and Urban Air Pollution.” NBER Working Paper #24183, Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w24183.

Grainger, Corbett A. 2012. “The distributional effects of pollution regulations: Do renters fully pay for cleaner air?” *Journal of Public Economics* 96:840–852.

Grainger, Corbett, Andrew Schreiber, and Wonjun Chang. 2018. “Do Regulators Strategically Avoid Pollution Hotspots when Siting Monitors? Evidence from Remote Sensing of Air Pollution.” Working paper.

Hu, Shishan, Scott Fruin, Kathleen Kozawa, Steve Mara, Suzanne E. Paulson, and Arthur M. Winer. 2009. “A wide area of air of air pollutant impact downwind of a freeway during pre-sunrise hours.” *Atmospheric Environment* 43 (16): 2541–49.

Lepeule, Johanna, Francine Laden, Douglas Dockery, and Joel Schwartz. 2012. “Chronic Exposure to Fine Particles and Mortality: An Extended Follow-up of the Harvard Six Cities Study from 1974 to 2009.” *Environmental Health Perspectives* 120 (7): 965–970.

Revesz, Richard L. 2015. *Environmental Law and Policy.* 3rd ed. St. Paul, MN: Foundation Press.

United States Environmental Protection Agency. 2012. *Regulatory Impact Analysis for the Proposed Revisions to the National Ambient Air Quality Standards for Particulate Matter.* Research Triangle Park, NC: Health, Environmental Impacts Division, Office of Air Quality Planning, and Standards, U.S. Environmental Protection Agency.

———. 2014. *Regulatory Impact Analysis for the Proposed Carbon Pollution Guidelines for Existing Power Plants and Emission Standards for Modified and Reconstructed Power Plants.* Research Triangle Park, NC: Health, Environmental Impacts Division, Office of Air Quality Planning, and Standards, U.S. Environmental Protection Agency.

———. 2015. *Regulatory Impact Analysis for the Clean Power Plan Final Rule.* Research Triangle Park, NC: Office of Air Quality Planning, Standards, Office of Air, and Radiation, U.S. Environmental Protection Agency.

United States Environmental Protection Agency. 2016. *Economic Analysis of the Formaldehyde Standards for Composite Wood Products Act Final Rule.* Washington, DC: Office of Chemical Safety and Pollution Prevention, U.S. Environmental Protection Agency. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2016-0461-0037.

Van Donkelaar, Aaron, Randall V. Martin, Michael Brauer, N. Christina Hsu, Ralph A. Kahn, Robert C. Levy, Alexei Lyapustin, Andrew M. Sayer, and David M. Winker. 2016. "Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors." *Environmental science & technology* 50 (7): 3762–3772.

Van Donkelaar, Aaron, Randall V. Martin, Robert JD Spurr, and Richard T. Burnett. 2015. "High-resolution satellite-derived PM2.5 from optimal estimation and geographically weighted regression over North America." *Environmental science & technology* 49 (17): 10482–10491.

Zou, Eric. 2018. "Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality." Working Paper. http://www.eric-zou.com/s/monitor_zou_201808.pdf.

Figures and Tables

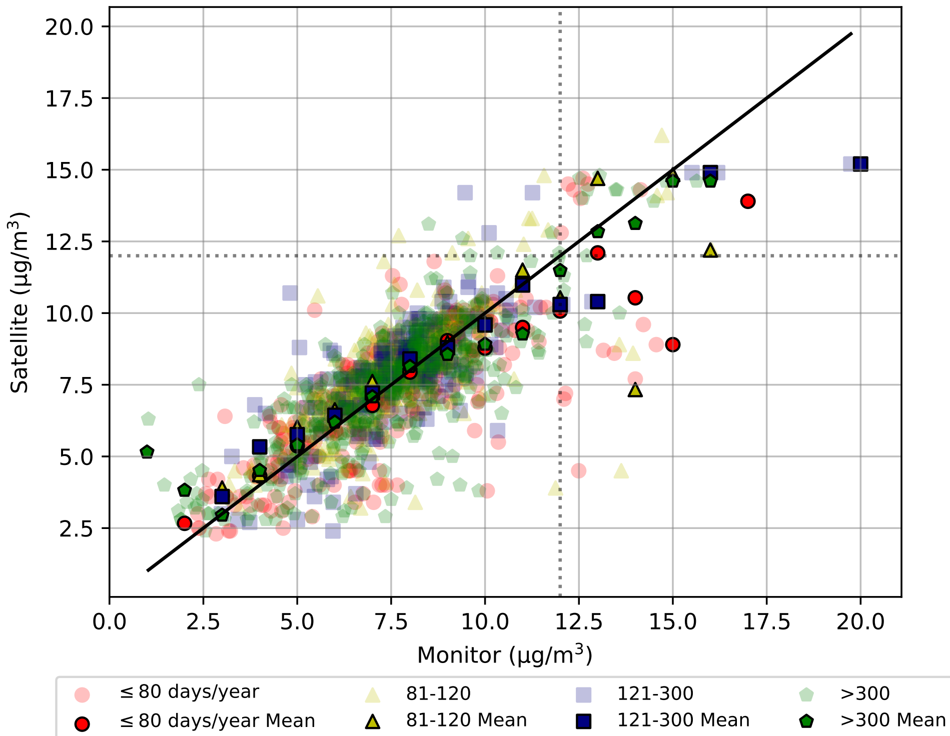

Figure 1: Satellite Readings versus Monitor Readings

Notes: Horizontal axis is annual average monitor reading. Vertical axis is the satellite-derived reading for the 0.01-by-0.01-degree cell where the monitor is located. Red circles indicate monitors that operate no more than 80 days per year; yellow indicate triangles 81–120 days; blue squares indicate 121–300 days; and green pentagons indicate more than 300 days per year. Faint markers indicate individual marker–grid cell pairs; bold markers indicate the average for every bin centered at integers on the horizontal axis, i.e., satellite average for monitor readings of 1 ± 0.5 µg/m^3. Dashed gray lines show the 12 µg/m^3 NAAQS threshold for nonattainment classification.

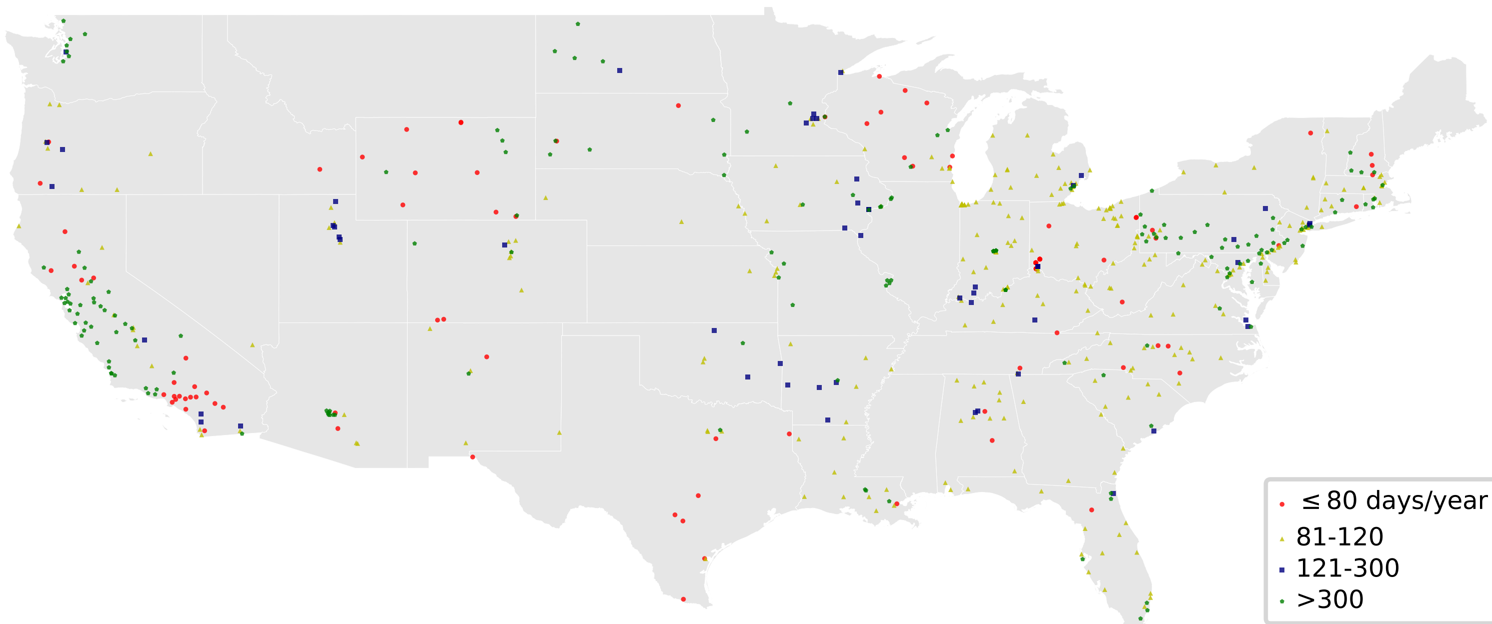

Figure 2: PM$_{2.5}$ Monitors by Temporal Coverage, 2011–2013
Notes: Categories determined using median of valid observation days from 2011 to 2013. Monitors must cover entire time period to be included in sample. Red dots denote monitors that operate no more than 80 days per year; yellow triangles denote 81–120 days per year; blue squares denote 121–300 days; and green pentagons denote at least 300 days per year.

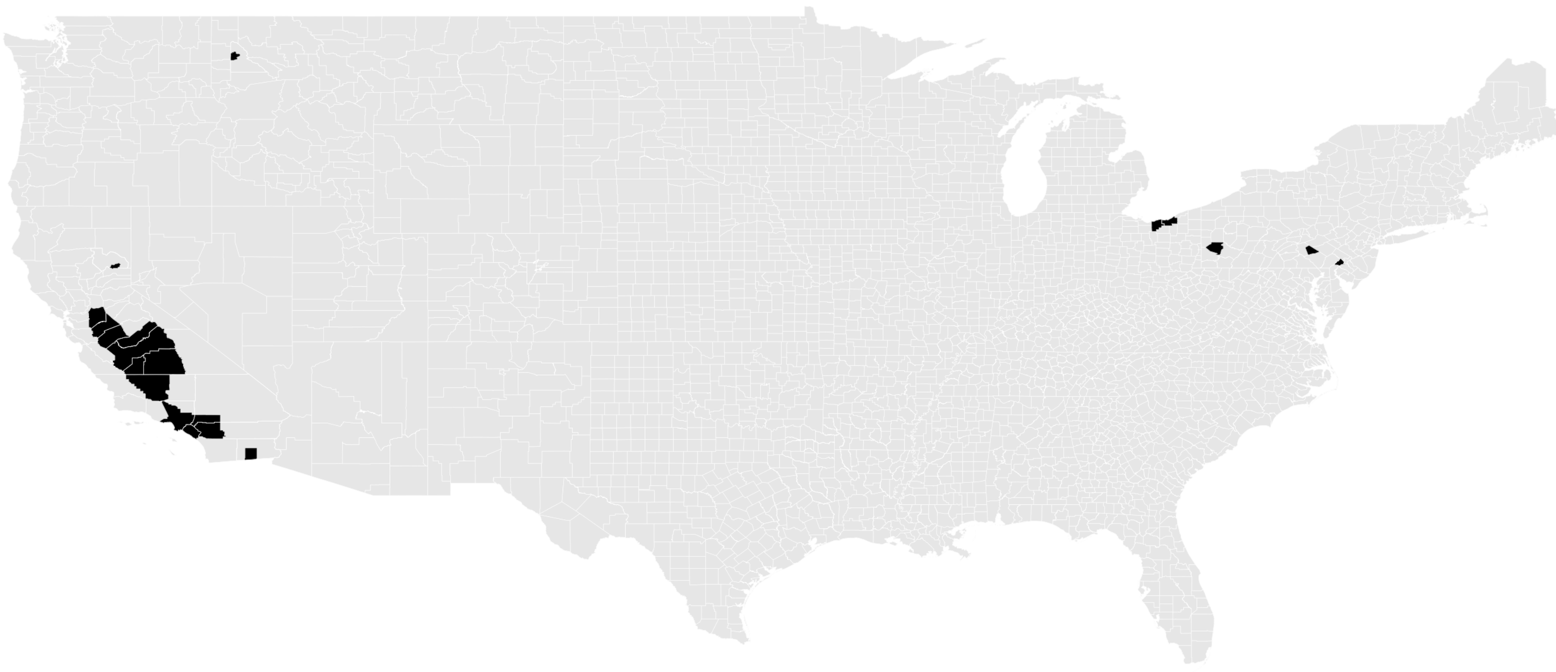

Figure 3: Clean Air Act Attainment Status, 2015
Notes: Darker areas are those classified as nonattainment with $PM_{2.5}$ 2012 primary standard of 12 $\mu g/m^3$.

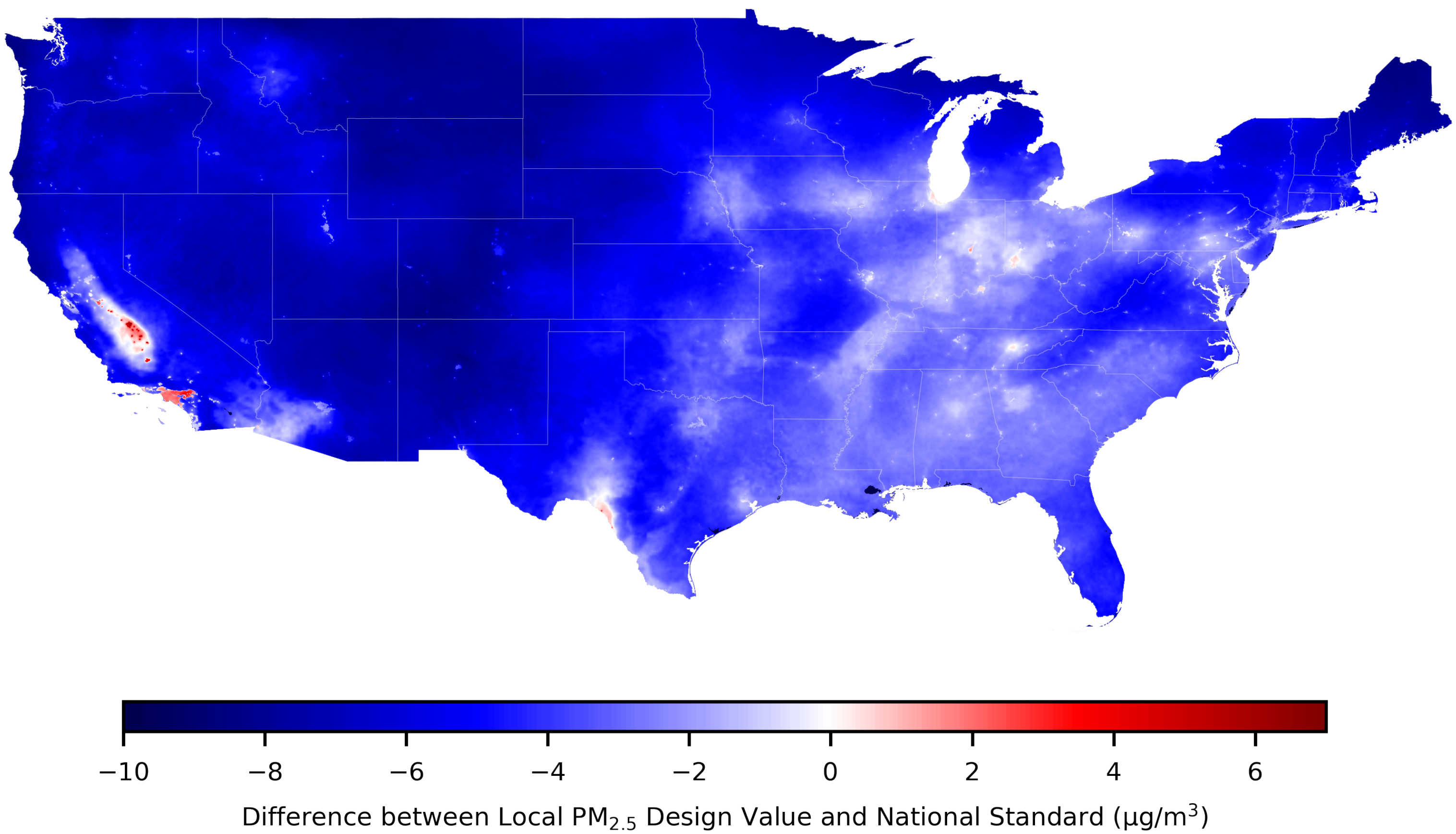

Figure 4: Satellite-measured $PM_{2.5}$ Design Values, 2015

Notes: $PM_{2.5}$ design values come from the satellite data described in Section 5. Plotted concentration is 3-year lagged average (2011–2013), which is the design value used to measure compliance with the NAAQS.

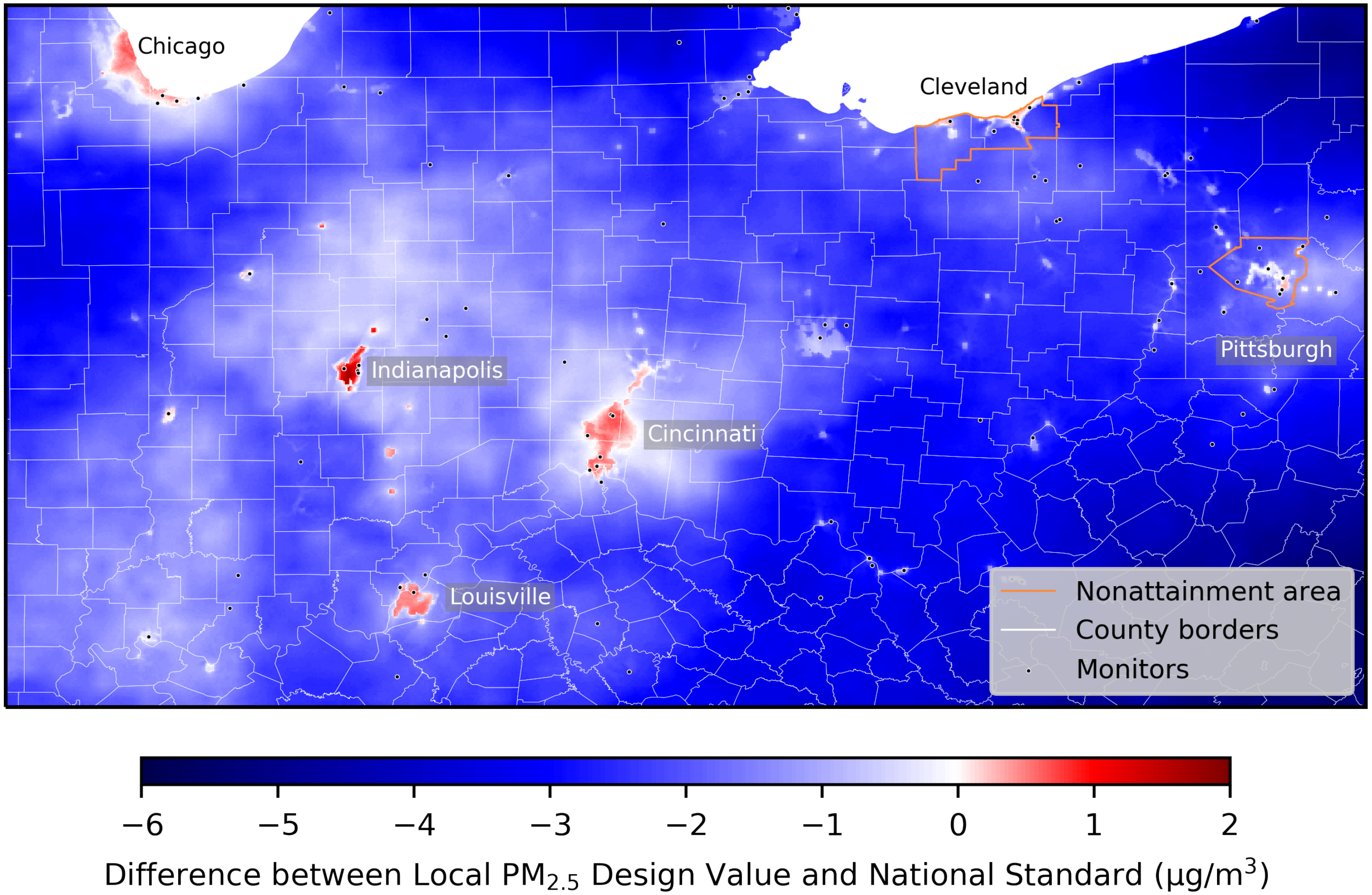

Figure 5: $PM_{2.5}$ Design Values and Attainment Status

Notes: Orange boundaries indicate official nonattainment areas for the $PM_{2.5}$ 2012 primary standard of 12 μg/m^3. Plotted $PM_{2.5}$ design values come from the satellite data described in Section 5 and are the average of years 2011–2013, the years of data that were used in making 2012 rule determinations. Monitor sample restricted to those used for NAAQS assessments.

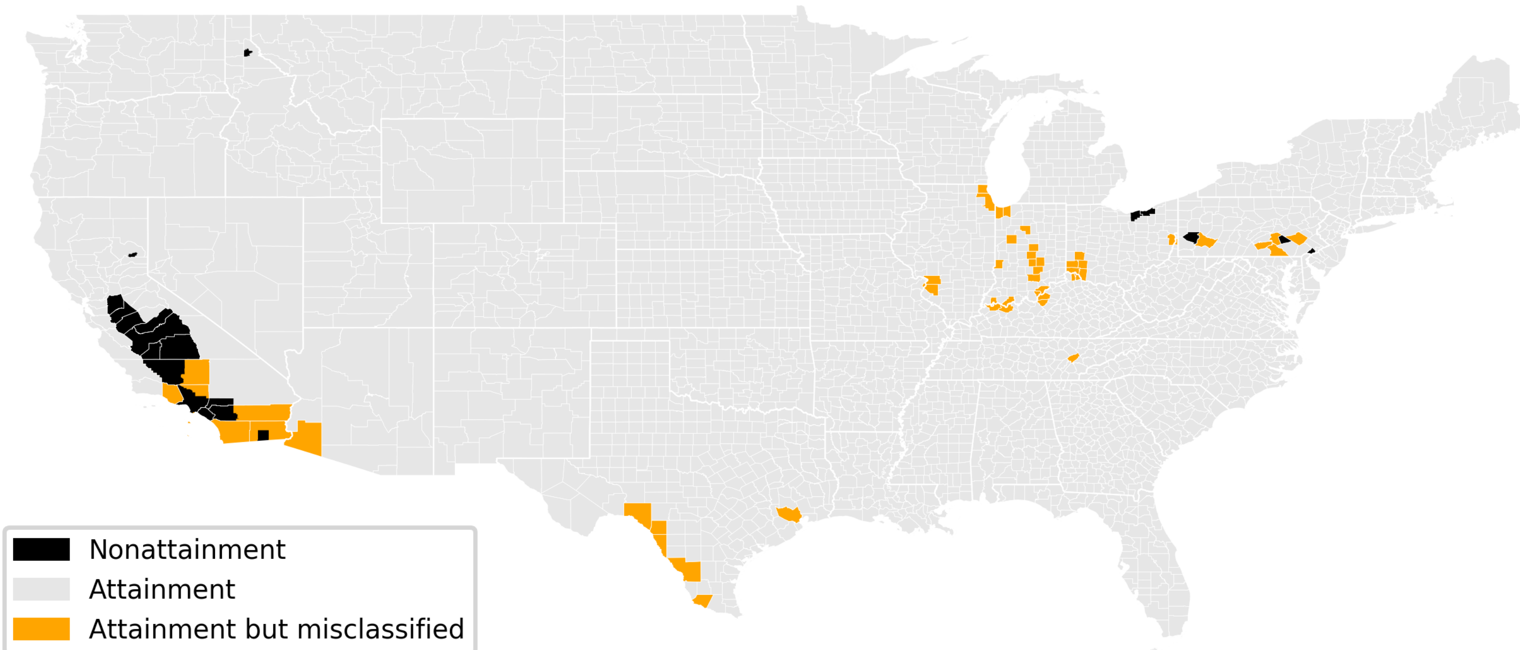

Figure 6: Areas Misclassified as Attainment for $PM_{2.5}$ Annual Standard

Notes: Black areas denote official nonattainment counties and sub-counties. Yellow areas are counties that are misclassified, i.e., counties that are officially attainment where the satellite data show that some portion of the county exceeds the NAAQS. Gray areas are correctly classified attainment counties.

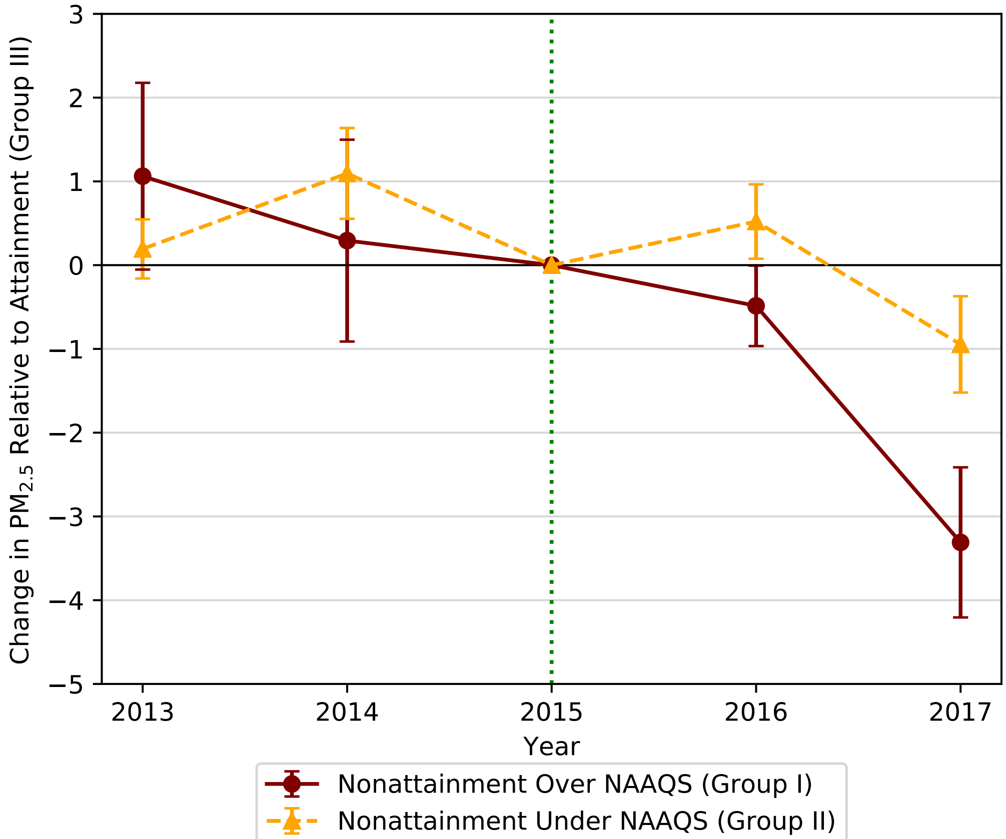

Figure 7: Event Study of Monitor Readings by Attainment Status and Group

Notes: Plot shows estimated coefficients from Equation (2). N=4,575. Number of monitors in Groups I, II, and III are 14, 49, and 852, respectively. Monitor sample includes only monitors used for NAAQS compliance and which began operation no later than 2013 and that operated through 2017. Red dots are the point estimates for monitors that were in nonattainment areas and that were higher than the NAAQS in 2015 (Group I in the text), i.e., the $\hat{\beta}_{1,y}$. Yellow dots are the point estimates for monitors that were in nonattainment areas and that were below the NAAQS in 2015 (Group II), i.e., the $\hat{\beta}_{2,y}$. The year 2015 is the reference year for both groups. Regressions also include monitor-level fixed effects. Bands around each dot show the 95 percent confidence interval. Standard errors clustered by monitor.

Table 1: $PM_{2.5}$ Monitor Counts by Frequency of Operation

	≤ 80 days	81–120 days	121–300 days	>300 days	Total
	A. Monitors Operating in the Given Year				
2010	121	345	87	148	701
2011	95	341	55	177	668
2012	106	296	93	202	697
2013	83	313	117	229	742
2014	118	319	77	264	778
2015	112	381	59	290	842
2016	99	330	131	326	886
	B. Monitors with 3 years of valid data for NAAQS assessment				
2013	72	274	73	157	576
2014	67	276	85	179	607
2015	90	290	62	215	657
2016	78	332	45	249	704
2017	77	299	96	274	746

Notes: Panel A reports the number of monitors designated as NAAQS primary compliance monitors which operated in the given year. Column 1 reports how many monitors operated no more than 80 days during that year, column 2 reports how many operated 81–120 days, and so on. Panel B reports how many monitors had sufficient data over the past three years that a design value could be calculated using that monitor. For example, a monitor that operated in 2016 but not 2015 would not be counted in 2016 while a monitor that operated in 2013–2015 would be counted in 2016.

Table 2: Misclassified Population by State

	Counties with no monitor	Counties with at least 1 monitor	Total
West Virginia	0	24,069	24,069
Tennessee	0	54,181	54,181
Arizona	0	195,751	195,751
Missouri	0	319,294	319,294
Kentucky	975,135	233,242	1,208,377
Pennsylvania	633,269	1,081,820	1,715,089
Ohio	945,497	1,240,213	2,185,710
Indiana	616,795	2,229,834	2,846,629
Texas	418,007	4,092,459	4,510,466
California	844,427	4,059,633	4,904,060
Illinois	6,437,475	0	6,437,475
Total	10,870,605	13,530,496	24,401,101

Notes: All misclassified counties in Illinois are counted as having no monitor because no monitor data were used in making attainment determinations in that state due to the monitors being deemed insufficiently accurate.

Table 3: Distribution of Demographic Groups Across Attainment Classifications

	(1)	(2)	(3)	(4)	(5)	(6)
	Percentage of Group Classified as			Nonattain. by Satellites	Monitors' False Negative Rate	Population
	Attainment	Misclassified	Nonattainment	(2) + (3)	(2) / (4)	(Millions)
Population	84.5	8.0	7.6	15.5	51.2	306.6
Rural	96.4	2.4	1.1	3.6	68.6	59.1
Urban	81.6	9.3	9.1	18.4	50.4	247.4
Race/Ethnicity						
White	88.7	6.7	4.6	11.3	59.6	196.0
Black	85.4	9.5	5.1	14.6	65.1	38.9
Hispanic	70.0	11.6	18.4	30.0	38.5	50.3
Asian	73.7	8.6	17.7	26.3	32.8	14.1
Other	72.4	10.1	17.5	27.6	36.5	30.9
Education						
No H.S. Diploma	81.0	8.6	10.5	19.0	45.0	29.8
H.S. Diploma	86.7	7.3	6.1	13.3	54.5	57.5
Some College	85.1	7.6	7.2	14.9	51.4	55.7
College Degree or More	84.7	8.1	7.1	15.3	53.3	55.3
Household Income						
<$35,000	86.0	7.6	6.3	14.0	54.6	38.8
$35,000–75,000	85.8	7.7	6.5	14.2	54.2	37.1
>$75,000	84.4	8.1	7.5	15.6	52.1	37.6

Notes: The first three columns of each row show show the percentage of the listed group that is attainment, misclassified, and nonattainment; e.g., 96.4 percent of people in rural areas are live in attainment counties. These columns sum to 100 for each row by construction. Data for population, share urban, and race/ethnicity come from 2010 census block-level counts. Data for income and education come from 2005–2010 ACS block group–level estimates. Education sample is people age 25 and older. Household income sample is by household; fourth column totals are number of households in the given income bin. NAAQS limit is from the 2012 $PM_{2.5}$ rule and is 12 $\mu g/m^3$.

Table 4: Effect of Nonattainment and NAAQS Status on Monitor Readings over Time

	(1)	(2)
Nonattainment×post	-1.0217***	
	(0.2134)	
Nonattainment×Over NAAQS×post		-2.3496***
		(0.4453)
Nonattainment×Under NAAQS×post		-0.6423***
		(0.2121)
2014	-0.1728***	-0.1728***
	(0.0432)	(0.0432)
2015	-0.4735***	-0.4735***
	(0.0430)	(0.0430)
2016	-1.1766***	-1.1766***
	(0.0508)	(0.0508)
2017	-1.4239***	-1.4239***
	(0.0606)	(0.0606)
R^2	0.828	0.829

Notes: N=4,575. Number of monitors in Groups I, II, and III are 14, 49, and 852, respectively. Outcome variable is annual average monitor reading of $\mu g/m^3$ $PM_{2.5}$. Monitor sample includes only monitors used for NAAQS compliance and which began operation no later than 2013 and that operated through 2017. The variable "post" is an indicator variable for years greater than 2015. "Over NAAQS" is an indicator variable for monitors whose annual average in 2015 exceeded the NAAQS limit of 12 $\mu g/m^3$. "Under NAAQS" is the complement of "Over NAAQS". Regressions also include monitor-level fixed effects. Standard errors clustered by monitor: ** p < .05, *** p< .01.

Supplementary Appendix

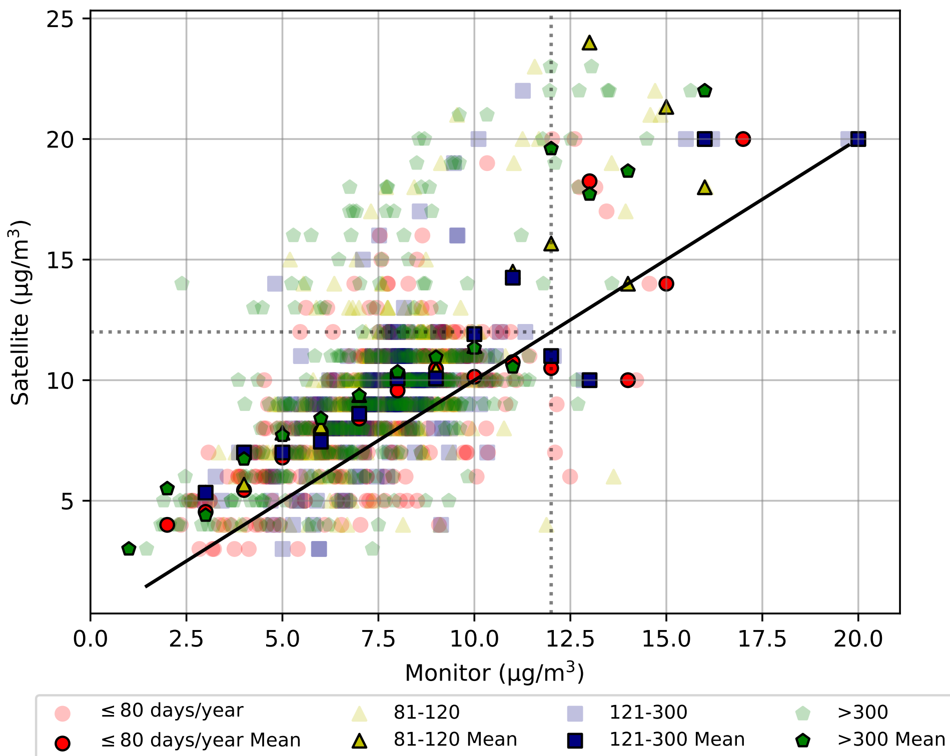

Figure A1: Globally Calibrated Satellite Readings versus Monitor Readings
Notes: Horizontal axis is annual average monitor reading. Vertical axis is the satellite-derived reading for the 0.01-by-0.01-degree cell where the monitor is located. Red circles indicate monitors that operate no more than 80 days per year; yellow indicate triangles 81–120 days; blue squares indicate 121–300 days; and green pentagons indicate more than 300 days per year. Faint markers indicate individual marker–grid cell pairs; bold markers indicate the average for every bin centered at integers on the horizontal axis, i.e., satellite average for monitor readings of 1 ± 0.5 µg/m^3. Dashed gray lines show the 12 µg/m^3 NAAQS threshold for nonattainment classification.

Table A1: Misclassified Counties and their Metro Areas

State	County	Core-based Statistical Area (Metro Area)
Arizona	Yuma County	Yuma, AZ
California	Imperial County*	El Centro, CA
	Kern County*	Bakersfield, CA
	Los Angeles County*	Los Angeles-Long Beach-Anaheim, CA
	Orange County*	Los Angeles-Long Beach-Anaheim, CA
	Riverside County*	Riverside-San Bernardino-Ontario, CA
	San Diego County	San Diego-Carlsbad, CA
	Ventura County	Oxnard-Thousand Oaks-Ventura, CA
Illinois	Cook County	Chicago-Naperville-Elgin, IL-IN-WI
	Lake County	Chicago-Naperville-Elgin, IL-IN-WI
	Madison County	St. Louis, MO-IL
	St. Clair County	St. Louis, MO-IL
Indiana	Bartholomew County	Columbus, IN
	Cass County	Logansport, IN
	Clark County	Louisville/Jefferson County, KY-IN
	Floyd County	Louisville/Jefferson County, KY-IN
	Hamilton County	Indianapolis-Carmel-Anderson, IN
	Jackson County	Seymour, IN
	Johnson County	Indianapolis-Carmel-Anderson, IN
	Lake County	Chicago-Naperville-Elgin, IL-IN-WI
	Marion County	Indianapolis-Carmel-Anderson, IN
	Porter County	Chicago-Naperville-Elgin, IL-IN-WI
	Shelby County	Indianapolis-Carmel-Anderson, IN
	Spencer County	
	Tippecanoe County	Lafayette-West Lafayette, IN
	Vanderburgh County	Evansville, IN-KY
	Vigo County	Terre Haute, IN
Kentucky	Bullitt County	Louisville/Jefferson County, KY-IN
	Campbell County	Cincinnati, OH-KY-IN
	Daviess County	Owensboro, KY
	Henderson County	Evansville, IN-KY
	Jefferson County	Louisville/Jefferson County, KY-IN
	Kenton County	Cincinnati, OH-KY-IN
Missouri	St. Louis city	St. Louis, MO-IL
Ohio	Butler County	Cincinnati, OH-KY-IN
	Clermont County	Cincinnati, OH-KY-IN
	Cuyahoga County*	Cleveland-Elyria, OH
	Hamilton County	Cincinnati, OH-KY-IN
	Jefferson County	Weirton-Steubenville, WV-OH
	Montgomery County	Dayton, OH
	Warren County	Cincinnati, OH-KY-IN
Pennsylvania	Berks County	Reading, PA
	Cumberland County	Harrisburg-Carlisle, PA
	Dauphin County	Harrisburg-Carlisle, PA
	Westmoreland County	Pittsburgh, PA
	York County	York-Hanover, PA
Tennessee	Roane County	Knoxville, TN
Texas	Harris County	Houston-The Woodlands-Sugar Land, TX
	Kinney County	
	Maverick County	Eagle Pass, TX
	Starr County	Rio Grande City, TX
	Val Verde County	Del Rio, TX
	Webb County	Laredo, TX
West Virginia	Brooke County	Weirton-Steubenville, WV-OH

* Partially misclassified county