A New Approach for Assessing Pollution-Reduction Benefits under the US Clean Air Act*

Joshua Murphy ® Eduardo Souza-Rodrigues ® Robert McMillan†

March 2026

Abstract

The 1970 Clean Air Act Amendments brought profound changes to environmental regulations in the United States, changes that have featured widely in subsequent empirical analyses of pollution. In this paper, we revisit both the pollution effects of local regulations introduced under the Act and the causal impacts of pollution on mortality and house prices; in turn, we propose a new approach for estimating those impacts. Drawing on data thought no longer to exist, we first show that the standard means of imputing regulatory status at the county level (in terms of the national air quality standards) is invalid: we set out an alternative measure of treatment based on the re-discovered official designations. Second, the non-random placement of monitors leads commonly-used estimators of the effects of the regulations on air pollution to deliver upward-biased estimates, which we correct using a balanced panel of monitoring stations. Third, we analyse the impacts of misclassification and measurement error when estimating the effects of air pollution, noting these can work in opposing directions. To resolve the ambiguity, our preferred two-stage least squares estimator of these effects uses a new instrument for treatment, corrects for the inflated first stage, and weights using population counts. The resulting estimates of the impact of air pollution on mortality and housing prices are larger and more precise than in well-known prior studies, notably Chay and Greenstone (2005). The evidence thus indicates that the literature has significantly *understated* the benefits of reducing particulates in this high-pollution era. Further, our estimates serve to rationalize more stringent air quality regulation in other high-pollution settings.

Keywords: air pollution, regulation, Clean Air Act, nonattainment, misclassification, monitoring, mortality, house prices, common trends, instrument validity

*We would like to thank Victor Aguirregabiria, Nate Baum-Snow, Karen Clay, Jonathan Hall, Ryan Kellogg, Kory Kroft, Nick Kuminoff, James Stock, Laura Turner, Matt Turner and seminar participants at Carnegie Mellon University and the University of Toronto for valuable comments and suggestions. Financial support is gratefully acknowledged from SSHRC and the University of Toronto Mississauga. Ben Lauzon provided excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of their respective organizations. All remaining errors are our own.

†Author names are in random order. Contact information: McMillan – University of Toronto and NBER, robert.mcmillan@utoronto.ca; Murphy – Finance Canada, Joshua.Murphy@fin.gc.ca; Souza-Rodrigues – University of Toronto, e.souzarodrigues@utoronto.ca.

1 Introduction

The Clean Air Act Amendments of 1970 had a profound impact on pollution regulation in the United States. Under the landmark legislation, nationally-uniform standards for ambient air quality were established, and regulators were required to identify so-called 'nonattainment' areas where the standards were not being met. Such areas then faced additional pressure to clean up source emissions within three years, unlike areas already deemed to be in compliance with the standards.[1] The substantial policy variation that resulted under the Act has been used widely in empirical analyses, as researchers have sought more credible identification when confronting pervasive endogeneity problems that arise in a pollution context.[2] In that vein, several well-known studies have estimated the benefits of reducing total suspended particulates ('TSPs') by exploiting regulatory variation induced by the Amendments, with estimated benefits covering a range of important outcomes: infant mortality (Chay and Greenstone, 2003), adult mortality (Chay, Dobkin, and Greenstone, 2003), housing values (Chay and Greenstone, 2005), fetal mortality (Sanders and Stoecker, 2015), and labor market earnings (Isen et al. (2017)).[3]

In this paper, we re-examine the effects of changes in local regulatory status under the Act on air pollution, and the causal impacts of pollution on mortality and house prices. While the Act has considerable historical importance, and ongoing research has sought to gauge impacts that emerged following its passage, some key features of the Act's implementation have remained obscure. By paying particular attention to the way the legislation was implemented in practice, we are able to shed new light on (i) the appropriate designation of treatment, and (ii) potential biases arising from the non-random rollout of the monitoring network under the Act. In turn, we propose a new approach to estimating the causal effects of pollution reduction in terms of mortality and house prices. The resulting estimates are larger and more precise than in prior work and, we argue, are relevant not only for our understanding of the Act's consequences, but also more generally, when formulating policy in other high-pollution settings today where regulations are often lacking.

We begin by considering the relationship between a county's initial nonattainment status and the change in its TSP concentration – the first stage in a popular instrumental variables approach used in the literature. The typical assumption has been that counties are the appropriate units for analyses contrasting regulatory stringency – Smith (2007) is an exception. In line with that assumption, the implicit notion has been that regulators assessed attainment status at the county level, comparing the highest-reading TSP monitor in each county to the national standards. On that basis, researchers have then contrasted the average change in TSP concentration in counties that initially barely met the national air quality standards with the same outcome in counties that initially barely failed.

[1] See Bachmann (2007) for relevant background.

[2] Not least, unobserved determinants of important health and housing outcomes may vary systematically across populations with different pollution exposures. Differences in the health behaviors and socioeconomic characteristics of populations in areas with high versus low levels of air pollution are documented in Graff Zivin and Neidell (2013).

[3] These studies all share an instrumental variables strategy based on a county's initial status (attainment versus nonattainment) with respect to the national standards, treating that as an instrument for the change in the county's TSP concentration.

The accuracy of county attainment status imputed in this way had not been examined prior to Murphy's 2017 doctoral thesis, as official data recording jurisdictions' regulatory status were thought to be no longer available. For example, Chay and Greenstone (2005) state that "The EPA did not begin to publicly release the annual list of TSPs nonattainment counties until 1978. We contacted the EPA but were informed that records from the early 1970s no longer exist."[4] Having located the official attainment/nonattainment designations, we shed fresh light on this issue, drawing on Murphy (2017). First, we document a frequent mismatch between the imputed versus the official designations. This arises because regulators based attainment status on monitors within an entire Air Quality Control Region (AQCR), not solely monitors within a county, as assumed in prior work.

This leads us to propose a new measure of treatment at the county level. In particular, we classify a county as being in nonattainment by interacting the official AQCR-level designation with the county-level imputation used in prior literature. In particular, a county is imputed as being in nonattainment by combining the official nonattainment designation initially defined at the AQCR level then merging to the county level. The resulting treatment applies to the year 1972.[5]

Using the official designations, we also show that the probability of being located in a nonattainment AQCR is similar among counties that barely met and barely failed the national TSP standard – that is, the regulatory discontinuity prior studies have intended to exploit is not present. The resulting misclassification thus invalidates the use of county-based regression-discontinuity designs; in addition, the underlying institutional arrangements do not support the use of the regression-discontinuity estimator in this setting.[6]

Next, we consider the role of monitor placement as part of the rapid expansion of the monitor network under the legislation. There is a presumption in prior studies that the growth and churning in the TSP monitor network around the time of the 1970 Amendments are not systematically biasing county-level trends in air pollution constructed from the full unbalanced monitor panel. Here we show that the design that compares changes in air pollution by initial attainment status, as implemented in the literature, overstates the effects of nonattainment on TSP reductions. Existing studies construct county-level trends in air pollution by averaging the daily TSP readings from all monitors operating in the same county during the same year. Each county's TSP trend is based on a potentially time-varying set of monitors, even though a balanced panel of counties is used in subsequent analyses. A county's air quality could thus appear to improve when in fact the decline in the average of its monitor readings from one year to the next was caused by: (i) removal of monitors

[4]Related, Sanders and Stoecker (2015) state that "Data on actual attainment status in the early 1970s are unavailable." Isen et al. (2017) state that "As highlighted in Chay and Greenstone (2003a, 2005), the EPA does not maintain historical records of actual county-level nonattainment status dating back prior to 1978."

[5]As described in more detail below, we impute the nonattainment designation for 1972 at the county level by identifying the monitor with the highest geometric mean of pollution readings and the monitor with the highest second-highest reading in 1970 within the county. If either the geometric mean exceeds 75 $\mu g/m^3$ or the second-highest reading exceeds 260 $\mu g/m^3$, the county is imputed as being a nonattainment county.

[6]The law required that plans designed to bring nonattainment areas into compliance should provide for the degree of emission reduction necessary to achieve the national standards, but not more. Hence, pressure to reduce source emissions increases with the degree of non-compliance, but may be negligible for areas just out of compliance.

with high concentration readings in an attempt to 'game' the national air quality standards; or (ii) placement of an additional monitor in a relatively unpolluted part of the county to satisfy the EPA's minimum network size requirements, which stimulated a rapid growth in the TSP monitor network in years following the 1970 Amendments.

We find that compositional changes such as these were important and occurred primarily in nonattainment counties. In turn, we construct more reliable air pollution trends for each county based on a fixed set of TSP monitors, active in both 1971 *and* the future years the trend is calculated over.

Comparing changes in air pollution by attainment status, we find that state plans to bring their nonattainment areas into compliance with the national air quality standards were implemented gradually, producing small but consistent year-on-year declines in air pollution from 1971 until the 1975 attainment deadline faced by most states. Estimates indicate that air pollution declined in nonattainment counties by less than is widely believed, especially during the first year the 1970 Amendments were in force.[7] Based on a comparison of estimates from balanced and unbalanced monitor panels, we find that prior studies overstate the effects of nonattainment on TSP reductions by around 40 percent. The first stage in the prior literature has thus been overstated.

Our main interest is in the causal impacts of air pollution on mortality and housing prices, leveraging the substantial variation in regulatory pressure induced by the Act. Bias concerns arise in this setting: as we show, the impacts of misclassification and measurement error when estimating the effects of air pollution can work in opposing directions. To resolve the ambiguity, we propose a new approach, central to which is an instrument for nonattainment, equal to one if a county is both imputed according to the treatment procedure described above and officially designated as nonattainment in 1972.

Using a long-difference design that is appropriate given the gradual rollout of state regulations, our 2SLS estimates indicate that air quality improvements caused substantial long-run reductions in mortality among the elderly, using a balanced monitor panel. Specifically, we find that a ten microgram reduction in mean TSP exposure between 1971 and 1977 caused a decrease of 89 deaths per 100,000 elderly residents over that time frame (statistically significant at a 5% level). In comparison, without correcting for non-random monitor placement and misclassification error, the benefits of air pollution reduction on elderly mortality would be significantly understated (67 deaths per 100,000). In terms of the total impact of the early 1970s USCAA policy on annual elderly mortality within non-attainment counties, our estimates indicate the policy led to a decrease of 7,712 annual elderly deaths between 1971-1977.

When considering impacts on housing values, our reassessment of the difference-in-differences estimator indicates that 2SLS estimates reported in Chay and Greenstone's classic (2005) study are over 45 percent too small, once a balanced monitor panel is used. That is, we find that a

[7]Specifically, after correcting for the misclassification and the non-random placement of monitors, we find counties that were designated as non-attainment in 1972 saw a decrease of 9.9 μg/m3 in mean TSPs from 1971 to 1975 relative to counties that were within attainment all else equal (statistically significant at a 1% level). In comparison, when using the incorrect imputed non-attainment measure, the decrease was 13.8 μg/m3.

reduction of 1 μg/m3 in air pollution causes an increase of 0.56% in the median property value (statistically significant at a 5% level), compared with a 0.38% increase in housing prices in Chay and Greenstone (2005) for the same reduction.

Based on the refinements we propose to existing designs using the 1970 Clean Air Act Amendments, these findings indicate that the leading estimates in the prior literature significantly *understate* the benefits of reducing particulate pollution. Such estimates matter when setting optimal policy – whether choosing among alternative air quality standards, caps on aggregate emissions, or taxes per unit of emissions. Of note, our larger estimates serve to rationalize more stringent air quality regulation in this and other high-pollution settings.

The rest of the paper is organized as follows: Section 2 describes the regulations in some detail. Section 3 sets out the data used. In Section 4, we highlight significant mismatches between the official Air Quality Control Region-level designations and the county-level designations used in the literature, before presenting our preferred treatment measure. In Section 5, we assess the role of the nonrandom rollout of the monitor network, documenting biases in difference-in-differences estimates arising from the changing composition of monitors within counties over time. Section 6 presents a formal discussion of competing biases in our setting, due to misclassification and mismeasurement. Section 7 then presents our preferred causal estimates of air pollution reductions on mortality, and Section 8, on house prices, based on the new instrumental variables approach we propose. Section 9 concludes.

2 Institutional Background

In this section, we provide relevant institutional background to the legislation, along with the associated regulatory history.

The Clean Air Amendments, enacted in December 1970, expanded the federal government's role in regulating air pollution significantly, contrasting with prior federal air pollution legislation, which had focused mainly on research, issuing grants, and offering collaborative mechanisms to support state-led efforts.[8] The Amendments required the federal Environmental Protection Agency (EPA) – itself set up under the Act in 1971 – to establish national ambient air quality standards (NAAQS) for key pollutants that were considered to have an adverse effect on public health, and also required states to submit plans specifying how the standards were to be achieved. We elaborate on these requirements in turn, before describing how the standards were applied in practice.

Air quality standards. The EPA promulgated national ambient air quality standards on April 31, 1971 for sulfur oxides, particulate matter, carbon monoxide, photochemical oxidants, hydrocarbons, and nitrogen dioxide. Reflecting the health risks posed by long-term exposure as well as short-term episodes of extreme air pollution, the national standards for TSP were formulated in terms of a *typical* reading and a *peak* reading. The typical reading was set at 75 $\mu g/m^3$, the annual

[8] Stern (1982) provides a thorough review.

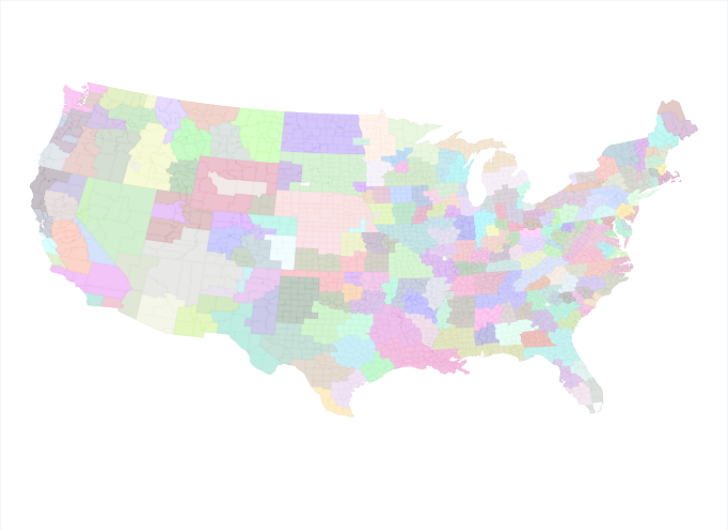


Figure 1: Map of the US showing coverage of the Air Quality Control Regions

geometric mean of 24-hour-averaged concentration readings not to be exceeded (the 'annual' standard). The peak reading was set to be 260 $\mu g/m^3$, the maximum 24-hour-averaged concentration not to be exceeded on more than one day per year (the 'short-term' standard).

State implementation plans. States were required to develop and submit plans to implement the standards within nine months, as per Section 110 of the Act, the requirements being further laid out in a Federal Register notice of August 14, 1971. Planning efforts were to focus on 'Air Quality Control Regions' (or AQCRs), designed to represent the broad atmospheric areas of the country. AQCRs originated in previous federal efforts to address interstate air pollution in the country's rapidly expanding metropolitan areas, which often extended across state borders. Many AQCRs also overlapped more than one state, their boundaries being based on urban-industrial concentrations, city or county lines, as well as considerations relating to climate, meteorology, and topography. Portions of states that were not already designated as AQCRs were grouped together to form new ones. Overall, there were 248 non-overlapping AQCRs, spanning the entire geographic area of the US – an order of magnitude fewer than the more than 3,100 counties (or their equivalents) in the US in 1970.

For each AQCR with pollution levels exceeding the national standards, state plans were to include a control strategy that set out the emission reductions needed to attain the standards within three years of the plan being approved by the EPA. Control strategies included a variety of possible measures – variously: emission intensity limits based on available control technologies; emissions charges, taxes or other incentives; closing or relocation of residential, commercial, or industrial facilities; changes to the schedules or methods of operation of commercial or industrial facilities; emissions controls on motor vehicles; traffic reduction measures; and the expansion of public transport. Measures outlined in the relevant control strategy were to be accompanied by legally enforceable compliance schedules. For the purposes of assessing the adequacy of the proposed

control strategies, the EPA required states to submit detailed emissions inventories and air quality data summaries, as well as estimates of the emissions and air quality levels expected to result from the application of the proposed measures.

Each state's plan had to provide for the establishment of an air quality surveillance system meeting minimum requirements for the number of monitoring sites in each AQCR, depending on the severity of the region's air quality problem and the region's population. At least one sampling site was to be placed in the area of estimated maximum pollutant concentration. States plans were to include a description of the existing and proposed surveillance systems and the rationale for the selection of sampling sites, following scientifically-based federal guidelines.[9] Because a comprehensive monitoring network did not yet exist, states had some discretion in terms of where monitors were placed, and also the total number in operation.

Attainment status designations. A Federal Register notice of May 31, 1972 listed the attainment status of each AQCR (and each state portion for interstate AQCRs) with respect to the standards, along with the region's attainment deadline. Of the 248 regions, 159 (65% of all AQCRs) had at least one state portion that did not meet the TSP standards, and most had attainment deadlines of three years for meeting the standards, by mid-1975.

The determination of each AQCR's attainment status involved close scrutiny of monitored or estimated air quality data, but was not necessarily based on well-defined or strictly adhered-to rules for comparing air quality data to the standards. Draft guidelines published by the EPA in October 1973 highlighted various points of interpretation that had frequently led to requests for further clarity. The guidelines sought "to serve as a useful step in the evolutionary development of a uniform and consistent set of criteria for relating ambient air quality data to the NAAQS" (EPA, 1973).

The first issue treated in the guidelines concerned the determination of attainment status in AQCRs with more than one monitor. The guidelines clarified that each monitoring site within an AQCR must meet the standard or the region would be considered in violation of the standard. Air quality improvements were of particular interest, and should therefore be monitored, in areas of maximum concentration and highest population exposure. A range of other issues were also addressed: significant digits; the definition of 'year' for annual standards (the calendar year); how to assess compliance against short-term standards that specified maximum concentrations not to be exceeded more than once per year in an AQCR with multiple monitors (each was allowed to register one violation).[10]

Nonattainment areas failing to reduce emissions so as to meet the standard could face a variety of penalties: the EPA could impose a federal plan; plants could potentially be closed or relocated;

[9] Some of the requirements and relevant considerations were outlined in "Guidelines: Air Quality Surveillance Networks," published by the EPA in 1971 (https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=9100K9L3.PDF).

[10] Further issues included: the number of years of 'clean' data required to demonstrate attainment; the influence of partial annual sampling on the probability of detecting violations of short-term standards; how to treat violations of the standards caused by recurring dust storms, forest fires, volcanic activity and other natural events; and whether all available air quality data should be used to determine compliance or only that from approved state air quality surveillance networks.

restrictions could be placed on industrial expansion or new facilities; and non-compliance could trigger legal enforcement actions.

1977 Amendments. In 1975, the EPA estimated (based on preliminary data) that most AQCRs – 132 of them – would likely not attain the TSP standards by their attainment deadlines. The EPA attributed this failure to "inadequate state plans, continued violations by a small number of major sources or numerous minor sources, windblown dust, and/or high background levels of the pollutant" (EPA, 1975).

In July 1976, the EPA called for revisions to state plans to address continued nonattainment problems, following a US Court of Appeals ruling in the case of *Union Electric v. EPA*, which found that the three-year deadline for attainment of the national standards could not be subject to delay on the basis of claims of technological or economic infeasibility. The revisions were to proceed in two waves, the first including all emissions limitations achievable with existing technologies (revised plans due to the EPA in July 1977), and the second, additional control measures that might be needed (revised plans due to the EPA in July 1978). In both instances, the revised plans had the potential to 'bite' – plant closures could be mandated, for example.

Congress enacted amendments to the legislation on August 7, 1977, acknowledging that the original timetables were too ambitious. The resulting 1977 Clean Air Act Amendments called for a fresh classification of areas by attainment status and for new state implementation plans to be submitted by January 1, 1979. The attainment date for the TSP standards was delayed to December 31, 1982. The Amendments also included an avenue for states to revise their attainment-nonattainment classifications periodically – if, for example, new nonattainment problems arose, unforeseen in the prior legislation.

The initial designations following the 1977 Amendments were published in a Federal Register notice on March 3, 1978. For TSPs, the designations varied at the county level or below to reflect "the spatially-limited nature of TSP violations," and rural areas experiencing violations of the standards due to windblown dust were able to claim attainment. Ultimately, ambient TSP concentrations did not improve much nationally in the latter half of the 1970s, after the first major push to implement the 1970 legislation but before implementation of the 1977 Amendments.

3 Data and Descriptive Statistics

In this section, we describe the data used in the analysis and provide some initial descriptive statistics. We also provide evidence of misclassifications in attainment status and the evolution of the composition of the network of monitors, with consequences for pollution measurements.

3.1 Data

Table 1: Summary Statistics

	Mean	SD	N
Imputed status	0.522	0.500	565
Official status	0.784	0.412	565
Interacted status	0.467	0.499	565
Geometric mean TSPs	75.255	43.197	565
First maximum value TSPs	224.946	209.030	565
Second maximum value TSPs	173.587	159.500	561
Log population	141.374	191.319	556
Population density	840.360	3659.579	558
Percent urban	64.491	26.698	558
Unemployment rate	4.603	1.813	558
Income	7378.483	1489.946	558
Percent poverty	11.642	6.709	558
Percent manufacturing	24.106	12.146	558
Percent white	90.260	11.657	558
Percent female	51.092	1.695	558
Percent senior	10.042	3.255	558
Percent high school plus	51.953	10.804	558
Percent college plus	10.162	4.636	558
Percent no plumbing	6.542	7.164	558
Percent owner occupied	66.987	9.089	558
Revenue	782.706	449.827	554
Property tax	332.978	159.402	554
General expenditures	670.318	223.639	554
Percent spend education	54.714	11.674	554
Percent spend highways	8.803	4.571	554
Percent spend welfare	4.590	5.664	554
Percent spend health	4.616	5.674	554
Percent spend education	56.642	52.945	56.246
Percent spend highways	9.550	8.117	9.722
Percent spend welfare	4.530	4.645	5.088
Percent spend health	4.376	4.836	4.973

TSP concentrations. Ambient air pollution measurements are from the EPA's national database, which stores readings from the universe of state, local, and federally-operated monitors throughout the United States. We obtain a *Quick Look Report* of annual statistics for all TSP monitors operating at any point between 1965 and 1975. Many of these monitors operate on the recommended once-every-six-days sampling schedule, producing approximately sixty 24-hour-averaged concentration readings per year. The annual statistics for each monitor include: the number of days monitored; the arithmetic mean of daily concentration readings; the geometric mean of daily concentration readings; and the second highest daily concentration reading. These data are used to

identify areas with initial pollutant concentrations above the national standards, and to construct measures of typical exposure to air pollution within a county over the course of a year.

Attainment Status. The official attainment status of each AQCR comes from the Federal Register notice of May 31, 1972. We assign the "*imputed*" attainment status to a county in 1972 whenever any of its TSP monitors exceeded either the annual or the short-term national standards in 1970, consistent with the classification used in the previous literature.

Outcomes and covariates. Mortality rates by age, sex, race, county, year and cause of death are from the 1968-78 Compressed Mortality File compiled by the National Center for Vital Statistics from the universe of death certificates filed in the United States. Housing data are from the same sources used by Chay and Greenstone: the 1972 and 1983 City and County Data Books for median housing values and other county-level characteristics.

4 Misclassification in Attainment Status

This section explores the issue of misclassified attainment status, in light of new information (described above) about the official designations.

Table 2 compares the actual attainment/nonattainment designations from the Federal Register against the imputed designations from the literature. The analysis is conducted at the county level, considering all 565 counties that were monitored for TSP in 1970. The table makes clear that the overall extent of misclassification is considerable, with almost 40 percent of counties (210 out of 565) being misclassified. Both types of error are present: 31 counties in AQCRs actually attaining the TSP standards (5.7 percent of the total) were predicted to be exceeding the standards; in contrast, 179 counties were predicted to be in attainment in AQCRs that were not attaining the TSP standards, or 31.5 percent of the total.

Table 2: Actual Versus Predicted Status of Counties

		Imputed Status Based on monitor data from 1970			
		Attainment	Nonattainment	Row Total	Percent Correct
Offical Status	Attainment	91	31	122	74.59
Federal Register	Nonattainment	179	264	443	59.59
May 31, 1972	Column Total	270	295	565	62.83

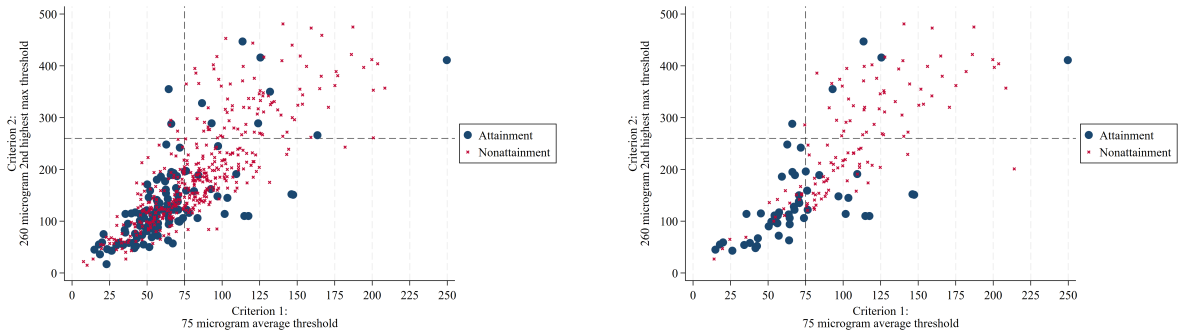
Table 3 shows that there has been substantial deviations from the TSP standard at the *AQCR level*, not just at the county-level: approximately 9% of the AQCR regions met the national standard but were classified as nonattainment region, and almost 10% exceeded the standards but were classified as being in attainment.[11]

[11] There are 194 observations in Table 3 rather than 248 because we do not have air pollution data in some AQCRs and so we are not able to create a predicted attainment measure for such regions.

Table 3: Actual Versus Predicted Status of AQCRs

		Imputed Status Based on monitor data from 1970			
		Attainment	Nonattainment	Row Total	Percent Correct
Official Status	Attainment	35	19	54	64.81
Federal Register	Nonattainment	17	123	140	87.86
May 31, 1972	Column Total	52	142	194	81.44

Figure 2 plots all combinations of the typical and the peak readings for each observation, along with the thresholds indicating the annual and short-term national standards: the scatterplot in panel (a) plots the observations at the county-level, while the scatterplot in panel (b) plots the AQCR-level data. In both panels, observations in attainment and nonattainment statuses are marked with dots and crosses, respectively. The two panels provide a better sense of the dispersion of the TSP measures around the thresholds. As is clear from the figure, there are attainment regions with TSP readings well above the threshold, and nonattainment regions with pollution levels well below the standard, providing further evidence that a strict rule was not followed in practice.



(a) County-level **(b)** AQCR-level

Figure 2: Official Attainment Designation in 1972 vs Air Quality Standard, County-level and AQCR-level

Table 4 below shows some county-level summary statistics separated by each of the four groups.

Table 4: Summary Statistics

	Imputed status		Official status		Interacted status	
	Attainment	Nonattainment	Attainment	Nonattainment	Attainment	Nonattainment
Number of counties	270	295	122	443	301	264
Geometric mean TSPs	51.421	97.069	63.775	78.416	56.811	96.284
First maxiumum value TSPs	141.842	301.007	201.545	231.39	166.961	291.056
Second maxiumum value TSPs	110.305	231.47	138.05	183.257	123.532	230.303
Log population	143.287	139.594	107.595	150.77	136.906	146.535
Population density	557.941	1103.235	313.669	986.195	524.07	1205.498
Percent urban	55.5	72.859	54.908	67.144	55.745	74.587
Unemployment rate	4.843	4.379	5.065	4.475	4.823	4.348
Income	7091.126	7645.953	6738.945	7555.563	7049.529	7758.239
Percent poverty	12.553	10.794	14.373	10.886	12.794	10.312
Percent manufacturing	22.634	25.477	19.562	25.365	22.367	26.114
Percent white	90.535	90.003	90.711	90.135	90.672	89.784
Percent female	50.864	51.304	50.854	51.158	50.884	51.332
Percent senior	10.322	9.782	10.926	9.798	10.445	9.578
Percent high school plus	51.694	52.194	50.817	52.267	51.589	52.373
Percent college plus	10.091	10.229	9.893	10.237	10.013	10.335
Percent no plumbing	7.695	5.469	8.38	6.033	7.707	5.198
Percent owner occupied	67.896	66.141	67.386	66.876	67.915	65.916
Revenue	769.258	795.038	784.117	782.316	764.203	803.782
Property tax	318.484	346.269	311.627	338.882	318.126	349.895
General expenditures	670.602	670.058	667.989	670.962	664.701	676.716
Percent spend education	56.642	52.945	56.246	54.29	56.679	52.476
Percent spend highways	9.55	8.117	9.722	8.549	9.662	7.824
Percent spend welfare	4.53	4.645	5.088	4.452	4.376	4.834
Percent spend health	4.376	4.836	4.973	4.517	4.418	4.842

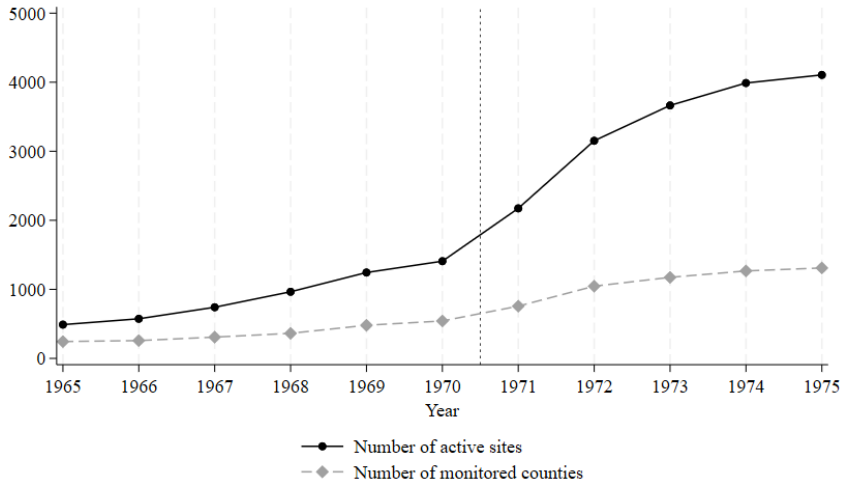
Building on the preceding evidence and looking ahead to the main empirical analysis, we propose a new treatment measure that operates at the county level but also reflects the official AQCR-level designation. Specifically, we classify a county as being in nonattainment by interacting the official AQCR-level designation, drawn from records thought no longer to exist, with a county-level imputation following the prior literature. In terms of the latter, to construct the county-level measure, we apply the national standard thresholds directly to county-level monitor data: a county is imputed as 'nonattainment' if any of its monitors records an annual geometric mean exceeding 75 $\mu g/m^3$ or a second-highest daily reading exceeding 260 $\mu g/m^3$. A county is then classified as 'nonattainment' under our new treatment measure if it satisfies both this county-level imputation and the official AQCR-level designation.

5 Systematic Changes in Monitor Composition

We now turn to the composition of the monitor network. We start with the evolution of the network, investigating the placement of new monitors and the retirement of operating monitors. Then we discuss the consequences of this evolution for the measurement of air pollution.

Evolution and Composition of the Network. Ambient TSP concentrations were already monitored widely in 1970, although truly national coverage was not attained until states were directed to bolster their monitoring programs as an integral part of their efforts to implement the national standards. Figure 3 shows the evolution of the number of monitors and number of monitored counties over time: it shows just how quickly the TSP monitor network was changing around the time of the 1970 Amendments. The fully implemented 1975 network comprised 4,331 monitors operating in 1,319 counties.

Figure 3: Growth of TSP monitor network



Although additions to the TSP monitor network after 1970 extended coverage to previously unmonitored counties, the majority of new stations were established in counties with at least one existing monitor.[12] The trend toward fuller coverage in monitored counties is highlighted in Table 5, which reports the number of stations operating in 407 counties that were continuously monitored for TSP between 1969 and 1974. In this balanced panel of counties, any increase in the number of stations is associated with an increase in the density of coverage. In counties with and without standards violations in 1970, the number of monitors roughly doubled between 1969 and 1974, even while several were retired. Such growth and churning opens the possibility that simple compositional changes might affect trends in the across-monitor average of daily TSP readings

[12]One reason for increasing the density of stations within a county relates to being better able to judge local population exposure. A significant share of the variation in monitor-specific annual arithmetic means in the fully implemented 1975 network – 45.1 percent – was within-county variation. This non-uniformity of TSP concentrations over small areas calls for spatially-detailed information.

taken within a county.

Table 5: Composition of Site Network

		Official Attainment				Official Nonattainment			
	Year	Total	Added this year	Removed this year	Per 100 square miles	Total	Added this year	Removed this year	Per 100 square miles
Imputed Attainment	1969	72	37	5	0.12	199	90	45	0.15
	1970	74	7	5	0.12	181	27	11	0.14
	1971	84	15	8	0.14	231	61	16	0.17
	1972	98	22	1	0.16	347	132	32	0.26
	1973	106	9	7	0.17	375	60	46	0.28
	1974	106	7	6	0.17	382	53	42	0.29
Imputed Non-attainment	1969	21	14	5	0.17	807	298	145	0.35
	1970	28	12	8	0.23	893	231	118	0.38
	1971	26	6	10	0.21	1182	407	139	0.51
	1972	38	22	3	0.31	1491	448	109	0.64
	1973	61	26	8	0.49	1663	281	124	0.71
	1974	60	7	8	0.48	1744	205	124	0.75

There are reasons to suggest that such changes in the monitor network were systematically related to attainment status. First, as documented in Table 5, counties with standards violations in 1970 were relatively well-monitored to begin with: in 1969, the number of stations per 100 square miles was 0.35 in nonattainment counties compared to 0.12 in attainment counties. If nonattainment counties were more likely to have monitors already placed in areas of maximum concentration, then they may have been forced to site new monitors in relatively unpolluted locations to satisfy the EPA's minimum network size requirements. Second, local regulators in nonattainment counties had strong incentives to remove monitors with high readings, as it would have allowed them to achieve the standards without undertaking costly action: the same incentives are not present in attainment counties.

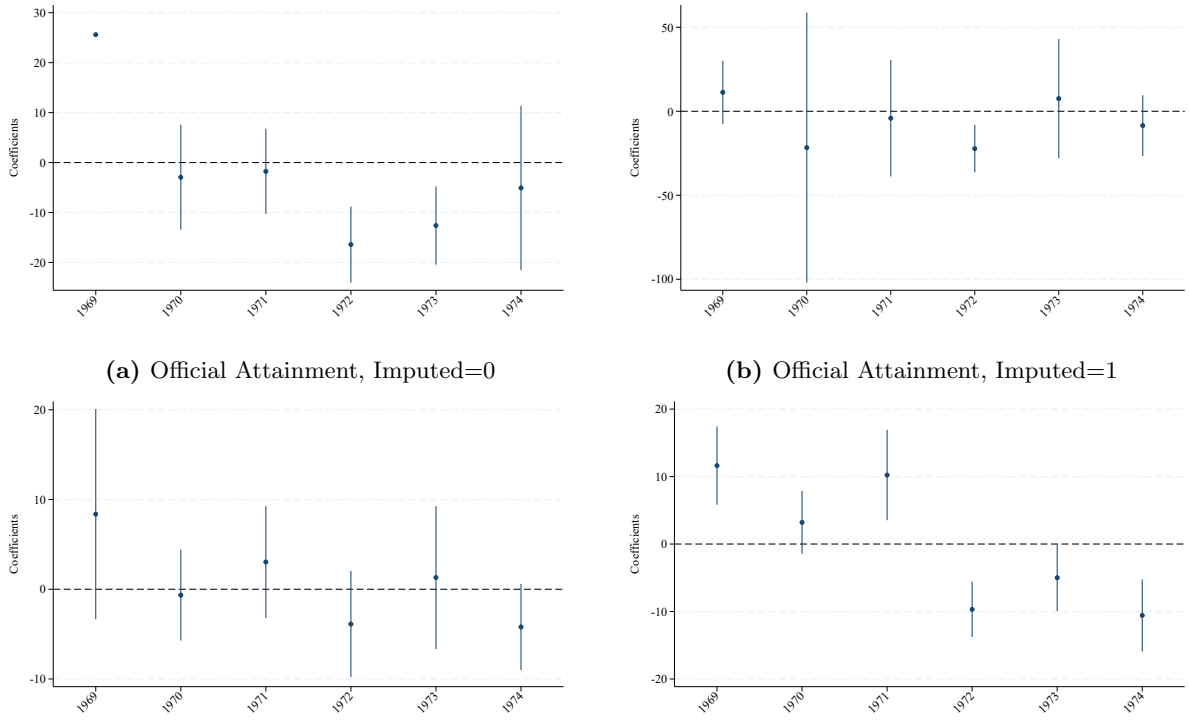
To test for differences in the pollution measurements of new and previously-established monitors operating in the same county and year, we regress a monitor's annual arithmetic mean of daily TSP readings on a dummy variable for non-attainment status (NA_{mt}), on a dummy variable for whether the monitor is new to the network (New_{mt}), on a third dummy variable for whether it was retired by the following year ($Removed_{mt}$), on the interaction between nonattainment status and New_{mt},

on the interaction between nonattainment status and $Removed_{mt}$, and on county-level fixed effects

$$\begin{aligned} TSP_{m,t}^{avg} &= \alpha_{c,t} + \alpha_1 NA_{mt} + \alpha_2 New_{mt} + \alpha_3 NA_{mt} \times New_{mt} \\ &\quad + \alpha_4 Removed_{mt} + \alpha_5 NA_{mt} \times Removed_{mt} + u_{mt}, \end{aligned} \tag{1}$$

where m indexes monitors, t indexes years, c indexes counties, and u_{mt} is an idiosyncratic shock. The coefficient α_1 captures the average differences in annual mean TSP concentration among regular monitors in nonattainment and attainment areas. The coefficient α_2 measures the average difference in pollution among new and regular monitors in attainment areas. The coefficient on the interaction term, α_3, captures the average differences in TSP concentration among new and previously-established monitors in nonattainment versus attainment areas – a difference-in-differences parameter. Coefficient α_4 measures the difference in average pollution among regular and to-be-removed monitors in attainment counties. And the last coefficient, α_5, is another difference-in-differences parameter, comparing regular and to-be-removed monitors in nonattainment versus attainment areas.

The parameters are estimated using ordinary least squares (OLS) for the period 1969–1974, based on monitors from the set of 407 continuously-monitored counties. The baseline group in the regression is composed of the "*regular*" monitors (not new, nor removed in the following year) in attainment areas.



(a) Official Attainment, Imputed=0

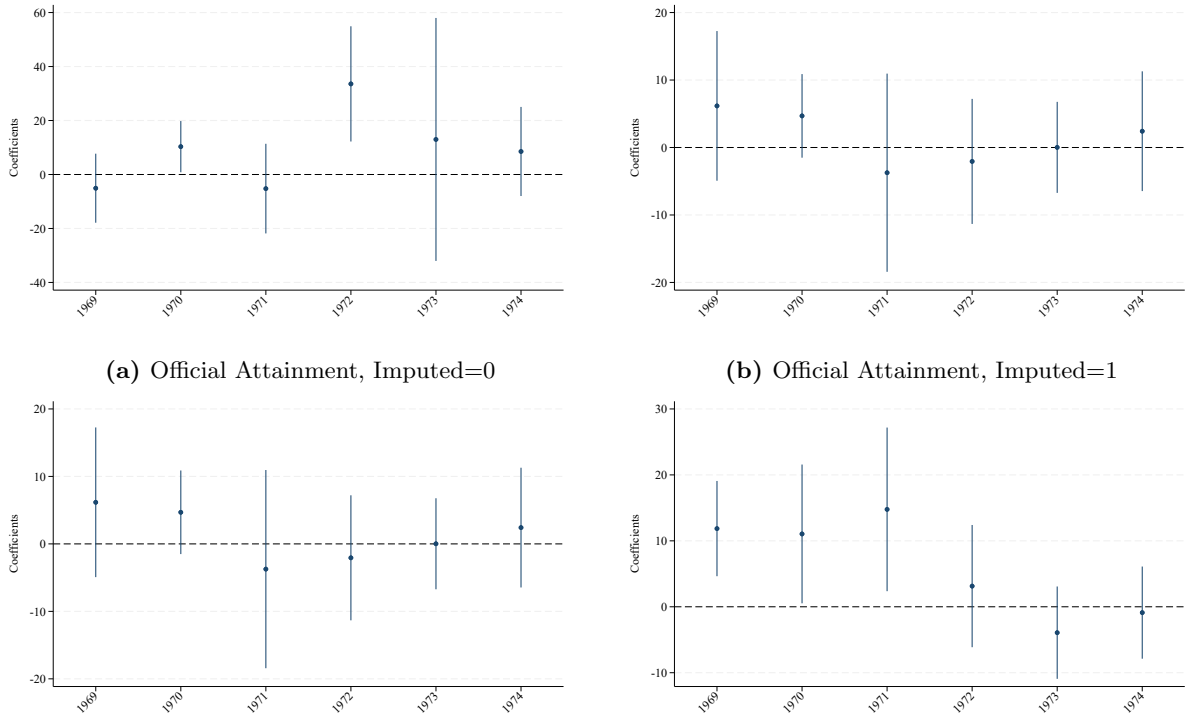
(b) Official Attainment, Imputed=1

(c) Official Nonattainment, Imputed=0

(d) Official Nonattainment, Imputed=1

Figure 4: Average TSP, New Sites

Figure 4 presents the results for the four groups. We focus initially on the interacted classification, presented in the third panel. As expected, regular monitors register higher levels of pollution on average in nonattainment areas than in attainment areas (though not significantly so). While new monitors in both attainment and nonattainment counties are placed, on average, in cleaner locations than the regular monitors, the differences are not statistically significant in both groups, and neither is the difference significant across groups. The removed monitors in attainment areas are also from cleaner areas. By contrast, in nonattainment counties, the removed monitors are from more polluted places. So, the evidence suggests that, while officials did not seem to 'game' the location of new monitors, they did so when retiring monitors in nonattainment counties with high levels of pollution.



(a) Official Attainment, Imputed=0

(b) Official Attainment, Imputed=1

(c) Official Nonattainment, Imputed=0

(d) Official Nonattainment, Imputed=1

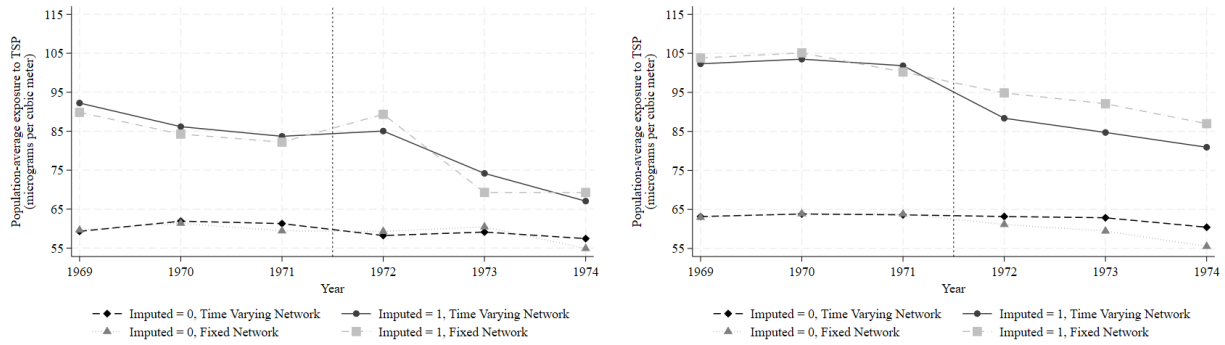
Figure 5: Average TSP, Removed Sites

Consequences for Pollution Measurements. Systematic changes in monitor composition occurred primarily in nonattainment counties, as we have just seen. Figure 6 allows one to compare the 'real' trend in population-average TSP exposure in nonattainment counties (based on a fixed set of monitors) with the 'biased' trend (based on a time-varying set of monitors).[13] The sample is limited to 181 nonattainment counties with at least one monitor that operated continuously between 1969 and 1974.

According to the trend constructed from a time-varying set of monitors, population-average

[13]Population-average TSP exposure in attainment counties was roughly constant over this period.

exposure to TSP in nonattainment counties declined sharply between 1971 and 1972, from 101 $\mu g/m^3$ to 88 $\mu g/m^3$, and eventually, to 81 $\mu g/m^3$ by 1974. These results are in line with the findings reported in the prior literature. Of note, however, according to the trend constructed from a fixed set of monitors, which is not affected by changes in the composition of monitors within nonattainment counties, population-average TSP exposure in nonattainment counties declined more gradually over the sample period, from 97 $\mu g/m^3$ in 1971 to 87 $\mu g/m^3$ in 1974 – half the decline observed in the 'biased' trend. Thus, while exposure to TSP fell significantly in nonattainment counties following the 1970 Amendments – by approximately 10 percent from 1971 levels by 1974 – the decline was much smaller than is widely believed.



(a) Attainment **(b)** Nonattainment

Figure 6: Air Pollution Over Time – Fixed vs Time-Varying Set of Sites

The evidence thus indicates that improvements in air quality were gradual, not immediate. Looking ahead, this will lead us to favor a 'long difference' design.

To quantify the importance of the resulting bias, we estimate the effects of nonattainment on the change in typical exposure to TSP within a county from 1971 to 1972 (and later years) using specifications from the literature, with one modification: we construct county-year TSP measures from a *fixed* set of monitors.

Table 6 reports output from OLS and WLS estimation of the following model for the change in typical exposure to TSP with county c between 1971 and $1971 + \ell$:

$$\overline{TSP}_{c,71+\ell} - \overline{TSP}_{c,71} = \kappa_\ell + \tau_\ell NA_{c,72} + \varepsilon_{c,71+\ell}, \tag{2}$$

where κ_ℓ is a constant and $NA_{c,72}$ is imputed nonattainment status in 1972 – an indicator for whether violations of the national TSP standards were recorded within the county in 1970. The coefficient of interest is τ, which measures the 'first-stage' effect of nonattainment on air pollution. The county-year TSP measures $\overline{TSP}_{c,71+\ell}$ and $\overline{TSP}_{c,71}$ used to produce estimates of τ_ℓ in the first three columns of Table 6 are constructed from a fixed set of monitors – those active in both 1971 and $1971 + \ell$. For comparison, the county-year TSP measures used to produce estimates of τ_ℓ in the last three columns are constructed from a time-varying set of monitors – those active in 1971, $1971 + \ell$, or both.

The simplest way to construct county-year TSP measures is to take the within-county average of monitor-specific annual arithmetic means. That is the approach taken in columns (1) and (4). Regardless of the length of the difference (*e.g.*, from 1971 to 1972 or from 1971 to 1974), estimates from a fixed set of monitors in column (1) are significantly smaller than estimates from a time-varying set of monitors in column (4). For example, according to the estimates in column (4), nonattainment caused a county's average TSP reading to decline by 8.34 $\mu g/m^3$ between 1971 and 1972. However, according to the estimates in column (1), nonattainment caused a county's average TSP reading to decline by only 3.12 $\mu g/m^3$ between 1971 and 1972 – less than half of the earlier estimate in column (4).

The county-year TSP measures used to produce estimates in columns (2) and (5) were constructed from observation-weighted averages, rather than simple averages, of monitor-specific annual means. Collapsing the monitor data to a county-year panel using observation weights has been the norm in prior studies because doing so is thought to reduce measurement error in each county's TSP measures – monitors that sample every day of the year produce less noisy estimates of actual air quality than monitors that sample infrequently. Were TSP concentrations homogeneous within counties, one would expect estimates in column 5 to be most precise because retaining all monitors and applying observation weights most effectively addresses measurement error under that assumption. Instead, estimates from the smaller, fixed set of monitors (columns 1 and 2) are more precise. Observation-weights have little effect on point estimates relative to the unweighted case.

Columns 3 and 6 report output from Weighted Least Squares estimation of equation 2, using county population size in 1970 as weights. All prior studies apply population weights because the health outcomes they examine are measured more precisely in more populous counties. Population weights increase the magnitude of the estimated effects and lower precision. With weighting, estimates from a fixed set of monitors remain half the size of estimates from a time-varying set of monitors. For example, estimates from a time-varying set of monitors in column 6 suggest that nonattainment caused population-average TSP exposure in nonattainment counties to fall by 15.97 $\mu g/m^3$ between 1971 and 1974. The corresponding estimate from a fixed set of monitors in column 3 is just 8.10 $\mu g/m^3$. If estimates from a fixed set of monitors are considered closer to the true effects, then these results suggest that difference-in-differences estimates from the literature overstate the effects of nonattainment on TSP reductions by a factor of two.

Table 6

	Coefficient on indicator for imputed nontattainment in 1972					
	Fixed set of sites			Time varying set of sites		
Dependent variable	No weights (1)	Site observation weights (2)	County population weights (3)	No weights (4)	Site observation weights (5)	County population weights (6)
Change 1971-1972	-3.115	-3.057	-4.102	-8.339	-7.769	-10.433
n=488	(1.307)	(1.244)	(1.943)	(2.146)	(1.960)	(2.453)
	[5.682]	[6.04]	[4.457]	[15.094]	[15.707]	[18.086]
Change 1971-1973	-7.530	-7.916	-8.222	-12.239	-11.991	-13.837
n=477	(1.521)	(1.518)	(2.810)	(2.278)	(2.106)	(2.738)
	[24.5]	[27.175]	[8.565]	[28.858]	[32.425]	[25.537]
Change 1971-1974	-5.303	-5.697	-8.104	-12.969	-12.993	-15.969
n=462	(1.607)	(1.594)	(2.784)	(2.395)	(2.250)	(3.021)
	[10.889]	[12.774]	[8.472]	[29.313]	[33.355]	[27.936]

Notes: F-statistic corresponding critical value 3.863. Each estimate derives from a separate regression. Heteroskedasticity-robust standard errors are given in parentheses, and F-statistics in square brackets. Sample size, referring to the number of counties with at least one monitor active in both years, is reported in the row heading. Imputed attainment status designations vary at the county level and are based on monitor data from 1970.

6 Econometric Framework

The section sets out a framework for understanding the potentially competing biases from misclassification and measurement error that arise in settings such as ours. We first lay out the general model, then obtain predictions that we apply to the current empirical analysis.

6.1 Setup: Model and Assumptions

We consider the classical linear IV model:

$$Y = \beta_0 + \beta_1 X^* + U,$$
$$X^* = \pi_0 + \pi_1 Z^* + V,$$

where $Z^* \in \{0, 1\}$ is a binary instrument. Let Z be the observed (possibly misclassified) version of Z^*, and X be the observed (possibly mismeasured) version of X^*. The structural and first-stage errors U and V are assumed to be mean zero and satisfy:

$$\mathbb{E}[Z^* U] = 0, \quad \mathbb{E}[Z^* V] \neq 0.$$

We allow for two forms of measurement issues:

- **Misclassification of the instrument:** Z is a misclassified version of Z^*:

$$q_0 = \mathbb{P}(Z = 1 \mid Z^* = 0) \quad \text{(false positive)},$$
$$q_1 = \mathbb{P}(Z = 0 \mid Z^* = 1) \quad \text{(false negative)}.$$

- **Measurement error in the endogenous regressor:** $X = X^* + \eta$, where η may be non-classical and correlated with Z or Z^*.

6.2 Standard IV Estimation without Measurement Error

In the absence of measurement error (i.e., $X = X^*$ and $Z = Z^*$), the IV estimator is given by:

$$\hat{\beta}_1^{IV} = \frac{\mathrm{Cov}(Y, Z)}{\mathrm{Cov}(X, Z)}.$$

Substituting $Y = \beta_0 + \beta_1 X + U$ and $X = \pi_0 + \pi_1 Z + V$ yields:

$$\mathrm{Cov}(Y, Z) = \mathrm{Cov}(\beta_1 X + U, Z) = \beta_1 \,\mathrm{Cov}(X, Z) + \mathrm{Cov}(U, Z).$$

Under the exclusion restriction $\mathbb{E}[ZU] = 0$, we have $\mathrm{Cov}(U, Z) = 0$, so:

$$\mathrm{Cov}(Y, Z) = \beta_1 \,\mathrm{Cov}(X, Z)$$
$$\Rightarrow \hat{\beta}_1^{IV} \xrightarrow{p} \beta_1.$$

Thus, the IV estimator is consistent under the standard assumptions.

6.3 Misclassification in the Instrument Only

6.3.1 First Stage Regression

We now suppose $Z \neq Z^*$ due to misclassification with probabilities q_0 and q_1.

We estimate the first stage:

$$X^* = \pi_0^{\text{naive}} + \pi_1^{\text{naive}} Z + \text{residual},$$

so:

$$\pi_1^{\text{naive}} = \frac{\text{Cov}(X^*, Z)}{\text{Var}(Z)}.$$

To compute $\text{Cov}(X^*, Z)$, apply the law of total covariance:

$$\text{Cov}(X^*, Z) = \mathbb{E}[\text{Cov}(X^*, Z \mid Z^*)] + \text{Cov}(\mathbb{E}[X^* \mid Z^*], \mathbb{E}[Z \mid Z^*]).$$

Term 1 (within-group covariance): Under the assumption that $Z \perp X^* \mid Z^*$ (i.e., Z is conditionally independent of X^* given Z^*), the first term is zero:

$$\text{Cov}(X^*, Z \mid Z^*) = 0.$$

Term 2 (between-group covariance): We compute:

$$\begin{aligned}
\mathbb{E}[X^* \mid Z^*] &= \pi_0 + \pi_1 Z^*, \\
\mathbb{E}[Z \mid Z^*] &= \mathbb{E}[Z \mid Z^* = 1] Z^* + \mathbb{E}[Z \mid Z^* = 0](1 - Z^*) \\
&= (1 - q_1) Z^* + q_0 (1 - Z^*).
\end{aligned}$$

Then:

$$\begin{aligned}
\text{Cov}(X^*, Z) &= \text{Cov}(\pi_0 + \pi_1 Z^*, (1 - q_1) Z^* + q_0 (1 - Z^*)) \\
&= \pi_1 \cdot \text{Cov}(Z^*, (1 - q_1) Z^* + q_0 (1 - Z^*)) \\
&= \pi_1 (1 - q_0 - q_1) \, \text{Var}(Z^*).
\end{aligned}$$

Misclassification attenuates the strength of the instrument. If $q_0 + q_1 \to 1$, the signal vanishes, and the first stage becomes weak.

6.3.2 Second Stage Estimator

To compute $\text{Cov}(Y, Z)$, again apply the law of total covariance:

$$\text{Cov}(Y, Z) = \mathbb{E}[\text{Cov}(Y, Z \mid Z^*)] + \text{Cov}(\mathbb{E}[Y \mid Z^*], \mathbb{E}[Z \mid Z^*]).$$

Term 1 (within-group covariance): Assuming $Z \perp Y \mid Z^*$ (i.e., Z is conditionally independent of Y given Z^*), and U is mean independent of Z^*, we have:

$$\text{Cov}(Y, Z \mid Z^*) = 0.$$

Term 2 (between-group covariance): We compute:

$$\begin{aligned}\mathbb{E}[Y \mid Z^*] &= \beta_0 + \beta_1(\pi_0 + \pi_1 Z^*),\\ \mathbb{E}[Z \mid Z^*] &= (1 - q_1)Z^* + q_0(1 - Z^*).\end{aligned}$$

So:

$$\begin{aligned}\text{Cov}(Y, Z) &= \text{Cov}(\beta_0 + \beta_1(\pi_0 + \pi_1 Z^*), (1 - q_1)Z^* + q_0(1 - Z^*))\\ &= \beta_1\pi_1(1 - q_0 - q_1)\,\text{Var}(Z^*).\end{aligned}$$

Then the IV estimator becomes:

$$\begin{aligned}\hat{\beta}_1^{IV} &= \frac{\text{Cov}(Y, Z)}{\text{Cov}(X^*, Z)}\\ &\xrightarrow{p} \frac{\beta_1\pi_1(1 - q_0 - q_1)\,\text{Var}(Z^*)}{\pi_1(1 - q_0 - q_1)\,\text{Var}(Z^*)} = \beta_1.\end{aligned}$$

So even with misclassification in Z, if X^* is observed and Z is used as an instrument, the IV estimator converges in probability to the true parameter β_1—but the instrument may become weak.

6.4 Adding Measurement Error in X

Now suppose $X = X^* + \eta$, where η may be correlated with Z.

6.4.1 First Stage: With Measurement Error and Misclassification

We now estimate:

$$X = \pi_0^{\text{naive}} + \pi_1^{\text{naive}} Z + \text{residual},$$

with:

$$\text{Cov}(X, Z) = \pi_1(1 - q_0 - q_1)\,\text{Var}(Z^*) + \text{Cov}(\eta, Z).$$

To see this, first note that

$$\mathrm{Cov}(X, Z) = \mathrm{Cov}(X^* + \eta, Z) = \mathrm{Cov}(X^*, Z) + \mathrm{Cov}(\eta, Z).$$

Consider first $\mathrm{Cov}(X^*, Z)$. Apply the law of total covariance:

$$\mathrm{Cov}(X^*, Z) = \mathbb{E}[\mathrm{Cov}(X^*, Z \mid Z^*)] + \mathrm{Cov}(\mathbb{E}[X^* \mid Z^*], \mathbb{E}[Z \mid Z^*]).$$

Term 1 (within-group covariance): Assume the following conditional independence:

$$Z \perp X^* \mid Z^* \quad \Rightarrow \quad \mathbb{E}[\mathrm{Cov}(X^*, Z \mid Z^*)] = 0.$$

This assumption means that conditional on the true instrument Z^*, the measurement error process that generates Z from Z^* is independent of the true endogenous regressor X^*. Formally, if the measurement error in the instrument is non-differential, meaning: $P(Z = 1|Z^*, X^*) = P(Z = 1|Z^*)$, then the assumption holds. This can happen when the measurement error mechanism is exogenous or randomized, or the data design creates a separation between the misclassification mechanism and the endogenous variable. In contrast, if places with higher X^* values are systematically more likely to misreport their Z^*, then the assumption fails.

Term 2: Compute via Conditional Expectations We know:

$$\begin{aligned}\mathbb{E}[X^* \mid Z^*] &= \pi_0 + \pi_1 Z^*, \\ \mathbb{E}[Z \mid Z^*] &= (1 - q_1)Z^* + q_0(1 - Z^*).\end{aligned}$$

Then:

$$\mathrm{Cov}(\mathbb{E}[X^* \mid Z^*], \mathbb{E}[Z \mid Z^*]) = \pi_1(1 - q_0 - q_1)\,\mathrm{Var}(Z^*).$$

So:

$$\mathrm{Cov}(X^*, Z) = \pi_1(1 - q_0 - q_1)\,\mathrm{Var}(Z^*).$$

And so

$$\mathrm{Cov}(X, Z) = \pi_1(1 - q_0 - q_1)\,\mathrm{Var}(Z^*) + \mathrm{Cov}(\eta, Z).$$

Next, we focus on the term $\mathrm{Cov}(\eta, Z)$. We find that:

$$\mathrm{Cov}(\eta, Z) = p_1 p_0(\mu_1 - \mu_0)(1 - q_1 - q_0),$$

where $p_1 = \mathbb{P}(Z^* = 1)$, $p_0 = 1 - p_1$, and $\mu_1 = \mathbb{E}[\eta \mid Z^* = 1]$, $\mu_0 = \mathbb{E}[\eta \mid Z^* = 0]$.

To see this, consider the decomposition from the law of total covariance:

$$\mathrm{Cov}(\eta, Z) = \mathbb{E}[\mathrm{Cov}(\eta, Z \mid Z^*)] + \mathrm{Cov}(\mathbb{E}[\eta \mid Z^*], \mathbb{E}[Z \mid Z^*]).$$

The first term, $\mathbb{E}[\text{Cov}(\eta, Z \mid Z^*)]$, equals zero under the assumption:

$$\eta \perp Z \mid Z^*,$$

that is, *the measurement error in X is conditionally independent of the observed instrument Z, given the true instrument Z^**. Here is why. By definition:

$$\text{Cov}(\eta, Z \mid Z^*) = \mathbb{E}[\eta Z \mid Z^*] - \mathbb{E}[\eta \mid Z^*] \cdot \mathbb{E}[Z \mid Z^*].$$

Under the conditional independence assumption $\eta \perp Z \mid Z^*$, we have:

$$\mathbb{E}[\eta Z \mid Z^*] = \mathbb{E}[\eta \mid Z^*] \cdot \mathbb{E}[Z \mid Z^*],$$

so:

$$\text{Cov}(\eta, Z \mid Z^*) = 0.$$

Taking expectations over Z^*:

$$\mathbb{E}[\text{Cov}(\eta, Z \mid Z^*)] = 0.$$

This assumption is justified if the measurement error in X and the misclassification in Z come from distinct processes that are independent once we condition on the true instrument Z^*. For example:

- η and Z arise from different sources (e.g., different survey sections or clerical mechanisms),
- The processes generating measurement error in X and misclassification in Z are unrelated once we know the true Z^*,
- There is no reverse causality or feedback from X^* into the misclassification mechanism of Z.

This simplifies our derivation of $\text{Cov}(\eta, Z)$ and focuses attention on the second term:

$$\text{Cov}(\mathbb{E}[\eta \mid Z^*], \mathbb{E}[Z \mid Z^*]).$$

We begin by using the law of total covariance:

$$\text{Cov}(\eta, Z) = \text{Cov}\left(\mathbb{E}[\eta \mid Z^*], \mathbb{E}[Z \mid Z^*]\right)$$

assuming $\eta \perp Z \mid Z^*$.

Next, write both conditional expectations as functions of Z^*:

$$\mathbb{E}[\eta \mid Z^*] = \mu_1 Z^* + \mu_0(1 - Z^*) = \mu_0 + (\mu_1 - \mu_0)Z^*,$$

$$\mathbb{E}[Z \mid Z^*] = (1 - q_1)Z^* + q_0(1 - Z^*) = q_0 + (1 - q_1 - q_0)Z^*.$$

Now compute the covariance of these two expressions, both of which are linear in Z^*:

$$\text{Cov}(\eta, Z) = \text{Cov}\left(\mu_0 + (\mu_1 - \mu_0)Z^*,\ q_0 + (1 - q_1 - q_0)Z^*\right)$$

$$= (\mu_1 - \mu_0)(1 - q_1 - q_0)\,\text{Var}(Z^*)$$

$$= (\mu_1 - \mu_0)(1 - q_1 - q_0)p_1p_0.$$

- The term $(\mu_1 - \mu_0)$ captures heterogeneity in the measurement error across true instrument status.
- $(1 - q_1 - q_0)$ reflects how informative the observed instrument Z is about the true Z^*. The more misclassification, the closer this term is to zero.
- p_1p_0 reflects the variation in the true instrument in the population; it's maximized when the distribution is balanced.

So the covariance between η and Z is stronger when:

- Measurement error depends significantly on the instrument ($\mu_1 \neq \mu_0$),
- The observed instrument is informative about the true one ($q_0 + q_1$ is small),
- And the true instrument is not rare in the population.

Putting these pieces together (noting that $\text{Var}(Z^*) = p_1p_0$):

$$\text{Cov}(X, Z) = \pi_1(1 - q_0 - q_1)p_1p_0 + (\mu_1 - \mu_0)(1 - q_0 - q_1)p_1p_0$$

Factoring common terms:

$$\text{Cov}(X, Z) = (1 - q_0 - q_1)\,p_1p_0\,(\pi_1 + \mu_1 - \mu_0)$$

Measurement error can strengthen or weaken the first stage depending on the sign of $\mu_1 - \mu_0$ and its interaction with the misclassification asymmetry.

- **First-stage strength** (π_1) captures how predictive the true instrument Z^* is for the true endogenous regressor X^*.
- **Measurement error heterogeneity** $(\mu_1 - \mu_0)$ captures how the mean of the measurement error in X varies with the true instrument Z^*. If $\mu_1 = \mu_0$, then η is mean independent of Z^*, and the second term drops out.
- **Misclassification severity** $(1 - q_0 - q_1)$ reflects how much information the observed instrument Z retains about the true instrument Z^*. When $q_0 + q_1 = 1$, the instrument becomes completely uninformative, and the covariance goes to zero.

- **Instrument variability** ($p_1 p_0$) is the variance of Z^* and is highest when the instrument is evenly split in the population. If Z^* is rare (e.g., $p_1 \to 0$), the signal shrinks.

Naive First-Stage Regression

The naive first-stage coefficient based on the observed (possibly misclassified) instrument Z is:

$$\pi_1^{\text{naive}} \xrightarrow{p} \frac{\text{Cov}(X, Z)}{\text{Var}(Z)}.$$

From previous derivations:

$$\text{Cov}(X, Z) = (1 - q_0 - q_1)\, p_1 p_0 \big(\pi_1 + \mu_1 - \mu_0\big),$$

and

$$\text{Var}(Z) = p_1 q_1 (1 - q_1) + p_0 q_0 (1 - q_0) + (1 - q_1 - q_0)^2\, p_1 p_0.$$

Thus:

$$\pi_1^{\text{naive}} \xrightarrow{p} \frac{(1 - q_0 - q_1)\, p_1 p_0 \big(\pi_1 + \mu_1 - \mu_0\big)}{p_1 q_1 (1 - q_1) + p_0 q_0 (1 - q_0) + (1 - q_1 - q_0)^2\, p_1 p_0}.$$

- **Misclassification effects (q_0, q_1):**
  - The numerator is scaled down by $(1 - q_0 - q_1)$. When misclassification is severe (i.e., $q_0 + q_1$ close to 1), the signal in $\text{Cov}(X, Z)$ shrinks toward zero.
  - The denominator is inflated by the added within-Z^* noise from misclassification, making the variance of Z larger.
  - Together, these effects always *weaken the apparent first stage*, since both a smaller numerator and a larger denominator reduce π_1^{naive}.
- **Measurement error heterogeneity in X ($\mu_1 - \mu_0$):**
  - If $\mu_1 > \mu_0$, the numerator is strengthened: the observed first stage can appear *stronger* than the true π_1.
  - If $\mu_1 < \mu_0$, the numerator is weakened further: the first stage is even smaller than with misclassification alone.
  - If $\mu_1 = \mu_0$, measurement error does not directly distort the first stage, and only misclassification matters.
- **Combined effect:**
  - Misclassification (q_0, q_1) always attenuates the IV, reducing its effective strength.
  - The heterogeneity in measurement error $(\mu_1 - \mu_0)$ can either counteract that attenuation (if positive) or exacerbate it (if negative).

6.4.2 Second Stage: Bias in IV Estimator

The IV estimator based on the observed regressor X and observed (possibly misclassified) instrument Z is:

$$\hat{\beta}_1^{IV} = \frac{\text{Cov}(Y, Z)}{\text{Cov}(X, Z)}$$

We assume:

- $Y = \beta_0 + \beta_1 X^* + U$
- $X = X^* + \eta$
- $X^* = \pi_0 + \pi_1 Z^* + V$
- Z is a misclassified version of binary Z^*, with misclassification probabilities $q_0 = \mathbb{P}(Z = 1 \mid Z^* = 0)$ and $q_1 = \mathbb{P}(Z = 0 \mid Z^* = 1)$
- $\mu_j = \mathbb{E}[\eta \mid Z^* = j]$ and $p_j = \mathbb{P}(Z^* = j)$ for $j \in \{0, 1\}$

Under standard IV assumptions (including $\mathbb{E}[U \mid Z^*] = 0$), we have:

$$\text{Cov}(Y, Z) = \beta_1 \,\text{Cov}(X^*, Z) = \beta_1 \pi_1 (1 - q_0 - q_1)\,\text{Var}(Z^*) = \beta_1 \pi_1 (1 - q_0 - q_1) p_1 p_0$$

From earlier derivations, we also have:

$$\text{Cov}(X, Z) = (1 - q_0 - q_1) p_1 p_0 \,(\pi_1 + \mu_1 - \mu_0)$$

Step-by-step Derivation

Plugging in the expressions:

$$\hat{\beta}_1^{IV} \xrightarrow{p} \frac{\beta_1 \pi_1 (1 - q_0 - q_1) p_1 p_0}{(1 - q_0 - q_1) p_1 p_0 \,(\pi_1 + \mu_1 - \mu_0)}$$

Canceling common terms:

$$\hat{\beta}_1^{IV} \xrightarrow{p} \frac{\beta_1 \pi_1}{\pi_1 + \mu_1 - \mu_0}$$

Dividing numerator and denominator by π_1:

$$\boxed{\hat{\beta}_1^{IV} \xrightarrow{p} \frac{\beta_1}{1 + \dfrac{\mu_1 - \mu_0}{\pi_1}}}$$

This expression highlights how measurement error and instrument misclassification distort the IV estimator:

- When $\mu_1 = \mu_0$, measurement error is mean independent of Z^*, and the estimator is consistent: $\hat{\beta}_1^{IV} \xrightarrow{p} \beta_1$.

- When $\mu_1 > \mu_0$, the denominator is inflated and the IV estimator is attenuated (downward biased).
- When $\mu_1 < \mu_0$, the denominator is deflated and the IV estimator is upward biased.
- The distortion is more severe when the measurement error heterogeneity $(\mu_1 - \mu_0)$ is large relative to the true first-stage strength π_1.

The full IV estimator becomes:

$$\begin{aligned}\hat{\beta}_1^{IV} &= \frac{\mathrm{Cov}(Y, Z)}{\mathrm{Cov}(X, Z)} \\ &= \frac{\beta_1 \pi_1 (1 - q_0 - q_1) \,\mathrm{Var}(Z^*)}{\pi_1 (1 - q_0 - q_1) \,\mathrm{Var}(Z^*) + \mathrm{Cov}(\eta, Z)}.\end{aligned}$$

Therefore:

$$\hat{\beta}_1^{IV} \xrightarrow{p} \frac{\beta_1}{1 + \dfrac{\mathrm{Cov}(\eta, Z)}{\pi_1 (1 - q_0 - q_1) \,\mathrm{Var}(Z^*)}}.$$

The estimator is attenuated if $\mathrm{Cov}(\eta, Z) > 0$, and inflated if it is negative. Even with a valid instrument Z^*, misclassification and heterogeneous measurement error can bias IV.

The full IV estimator becomes:

$$\begin{aligned}\hat{\beta}_1^{IV} &= \frac{\mathrm{Cov}(Y, Z)}{\mathrm{Cov}(X, Z)} \\ &= \frac{\beta_1 \pi_1 (1 - q_0 - q_1) \,\mathrm{Var}(Z^*)}{(1 - q_0 - q_1) \left[\pi_1 \,\mathrm{Var}(Z^*) + (\mu_1 - \mu_0) p_1 p_0\right]} \\ &= \frac{\beta_1}{1 + \dfrac{(\mu_1 - \mu_0) p_1 p_0}{\pi_1 \,\mathrm{Var}(Z^*)}}.\end{aligned}$$

This expression can be interpreted directly in terms of the misclassification probabilities q_0, q_1, the first-stage strength π_1, and the conditional means of the measurement error:

$$\hat{\beta}_1^{IV} \xrightarrow{p} \frac{\beta_1 \pi_1 (1 - q_0 - q_1) \,\mathrm{Var}(Z^*)}{\pi_1 (1 - q_0 - q_1) \,\mathrm{Var}(Z^*) + (\mu_1 - \mu_0)(1 - q_0 - q_1) p_1 p_0}.$$

This formulation highlights that the IV estimator's probability limit is distorted by the difference in measurement error means across true instrument groups $(\mu_1 - \mu_0)$, scaled by their population shares $(p_1 p_0)$, and by the degree of instrument misclassification $(1 - q_0 - q_1)$. When $(\mu_1 - \mu_0) > 0$, the denominator inflates and the estimator is attenuated; when $(\mu_1 - \mu_0) < 0$, the estimator may be biased upward. The impact is more severe when p_1 and p_0 are both sizable (i.e., the instrument is balanced in the population).

6.5 Relevance to Our Application

We now relate the preceding general discussion to important features of our application.

First, the term $\mu_1 - \mu_0$ is the differential measurement error in pollution among the *true* treatment versus control group. Specifically, let $Z*$ be the true treatment status (in our application, the interacted status). Then

$$\mu_j = E(\eta | Z* = j) = E(X - X* | Z* = j),$$

Next, X is the mismeasured pollution (based on the unbalance panel of monitors) and $X*$ is the correctly measured pollution (based on the balanced set of monitors). So, if there is no measurement error, then $\mu_j = 0$, for both groups. Our claim is that $\mu_0 = 0$, as we see evidence of gaming only in the true treatment group, not in the control group. For the true treated group, $\mu_1 ! = 0$.

In the panel data, X and $X*$ are actually first differences, $\Delta X = X_{t+1} - X_t$, so we are really talking about mismeasured differences over time. So

$$\mu_j = E(\Delta X - \Delta X* | Z* = j).$$

For $Z* = 1$, observed pollution declined a lot, so ΔX is large and negative, while true pollution has not declined as much, so ΔX is small negative. This implies $\mu_1 = E(\Delta X - \Delta X* | Z* = 1) < 0$.

In the first stage, we expect $\pi_1 < 0$ (treatment status pushes pollution down), and $\mu_1 - \mu_0 = \mu_1 < 0$. So differential measurement error across groups makes the first stage appear even more negative, giving the (wrong) impression of a strong IV. This goes against the misclassification problem, which gives the wrong impression of a weak IV, but it dominates in our health application. So correcting the false impression of strong IV makes the correct IV weaker.

For the second stage, misclassifications do not matter – in the worst case, think of flipping classifications entirely, which would amount to just relabelling Z, with no impact on the estimated coefficient of interest. But $\mu_1 - \mu_0$ having the same sign as $\pi (< 0)$ wrongly pushes the estimate downward. Once we correct the differential measurement error, we obtain a bigger, and now-correct, β.

Having considered biased, we now turn to precision: the correction leads to weak IV which will lead to a less precisely estimated β.

7 Re-assessing Impacts of Pollution

We now turn to the main estimation exercise in the paper, applying our approach to generate new estimates of the benefits of pollution reduction in terms of mortality rates and housing prices.

7.1 Impacts on Mortality Rates

First we discuss the long-run effects of the 1970 Clean Air Amendments on adult mortality. Table 7 illustrates the long-run impacts of air pollution reduction on the elderly mortality rate. In the first stage, we look at the impact of county-level non-attainment designation on the change in mean TSP level between 1971 and 1975. All six specifications utilize the change in the elderly mortality rate between 1971 and 1977 as the second stage outcome. In both first and second stages, we control for sociodemographic characteristics drawn from the Census, including the unemployment rate, education, population density, as well as percentage of labour force in manufacturing jobs, and region fixed effects. In each column of table 7, we analyze how different choices of instrument (imputed, official or interacted designations) and type of monitoring network (fixed or time varying) affect both the first and second stage estimates.

Focusing on the first stage results, we show clear evidence that failing to account for the non-random placement of the site network systematically inflates the first-stage estimates regardless of the choice of instrument. For example, when using the imputed non-attainment instrument without correcting for the churn in monitoring sites, we find that counties that were imputed as non-attainment saw, on average, a statistically significant decrease of 13.87 $\mu g/m^3$ in their average TSP levels. However, when utilizing a balanced panel of monitoring sites, this effect decreases by over 25% to 10.38 $\mu g/m^3$ but remains statistically significant. This pattern is also visible across other specifications.

We have already discussed the implications of an inflated first-stage on the second-stage, and our results conform with that. That is, we find that the second-stage estimates were attenuated in all cases when using the time-varying site network relative to those using the balanced panel of sites. Finally, we interact both the official and imputed designations as an instrument to address misclassification and measurement errors. After correcting both sources of bias, we find that a reduction of $10\mu g/m^3$ in average TSP caused a statistically significant cumulative reduction of 8.9 deaths per 100,000 elderly residents between 1971-1977 on average. We note that our results differ significantly from those found by Chay et al. (2003) for two reasons. First, their difference-in-differences design was estimated over a short horizon, during which the policy's effect on air pollution may not have yet materialized in a significant way. Second, using the imputed non-attainment measure as an instrument and a time-varying monitoring network likely led to an inflated first stage while attenuating the second stage estimates.

Table 8 shows the long-run effects of the policy on different types of mortality rates across different age groups utilizing our preferred interacted instrument and the fixed site network. We find, unsurprisingly, that elderly residents are the ones who benefit the most from air pollution reductions. Across other age groups, our estimates are not statistically significant.

Table 7

	A. First stage (Change in mean TSPs in $\mu g/m^3$)					
	(1)	(2)	(3)	(4)	(5)	(6)
Pt. Estimate	−13.869	−10.384	−9.468	−4.215	−13.003	−9.993
Std. Error	(3.076)	(2.49)	(5.086)	(3.152)	(2.983)	(2.434)
Effective F-Stat	20.332	17.392	3.466	1.788	19.001	16.861
	B. 2SLS (Change in mortality per 10,000 residents)					
Pt. Estimate	0.668	0.893	0.553	1.243	0.689	0.897
Wald CI	[0.14, 1.20]	[0.12, 1.67]	[−0.49, 1.60]	[−1.38, 3.86]	[0.13, 1.25]	[0.10, 1.70]
AR CI	[0.19, 1.36]	[0.22, 2.01]	[., .]	[., .]	[0.18, 1.43]	[0.20, 2.04]
Nonattainment	Imputed	Imputed	Official	Official	Interacted	Interacted
Fixed monitor network	No	Yes	No	Yes	No	Yes

7.1.1 Valuing the Benefits of the 1970 CAA on Mortality Rates

Table 8

	Mortality rates (deaths per 10,000 residents)			Annual mortality (# deaths)
Age Group	All cause mortality (1)	Cardioresp. mortality (2)	Non-cardioresp. mortality (3)	Column (1) × first stage impact × nonattain. pop. (4)
25 to 44 years	0.083	0.039	0.044	-2095
AR CI	[-0.007, 0.227]	[-0.000, 0.097]	[-0.023, 0.144]	
45 to 64 years	0.137	0.128	0.009	-2647
AR CI	[-0.082, 0.489]	[-0.001, 0.335]	[-0.140, 0.208]	
65 to 84 years	0.897	0.725	0.172	-7712
AR CI	[00.202, 2.044]	[00.138, 1.771]	[-0.159, 0.527]	
$\geq$ 85 years	-1.253	-1.673	0.419	996
AR CI	[-5.058, 1.907]	[-5.083, 1.159]	[-1.006, 1.793]	

We now discuss the welfare implications of these results. Our results imply that the 1970 Clean Air Amendments saved, on average, 7,712 elderly residents every year over that period of time. Adding up the reductions in mortality across all age groups of individuals aged 84 or less, we estimate that the mortality benefits of the policy were north of 87 billion dollars[14]. Not correcting for both sources of bias would have understated the policy's benefits by nearly 24 billion dollars. In other words, the policy's true benefits with respect to mortality would have been underestimated by over 27%. This finding indicates that the significant benefits of more stringent air pollution regulation have been understated in prior research.

7.2 Impacts on Housing Prices

Next, we conduct a similar exercise using housing prices as our outcome of interest. Table 9 compares first and second stage results across different choices of instruments and site network. The specification illustrated in table 9 looks at the effects of reductions in air pollution between 1970 and 1980 on changes in median house value. We include controls for various socioeconomic and demographic covariates in 1970. Similar to our mortality results, we find an inflated first stage when we do not account for the non-random placement of sites. The consequences on the second stage are the same as in the case of mortality: the second stage estimates are attenuated. After correcting for the misclassification error and the time-varying site network, we find that a reduction of 1 μg/m^3 caused, on average, a statistically significant increase of 0.78% all else equal.

Table 9

	(1)	(2)	(3)	(4)	(5)	(6)
	A. First stage (Change in mean TSPs (1/100) in $\mu g/m^3$)					
Pt. Estimate	−0.076	−0.046	−0.048	−0.038	−0.073	−0.047
Std. Error	(0.02)	(0.015)	(0.016)	(0.012)	(0.022)	(0.015)
Effective F-Stat	13.88	10.09	9.48	9.27	10.58	9.61
	B. 2SLS (Change in median log house value)					
Pt. Estimate	−0.513	−0.841	−0.554	−0.703	−0.508	−0.781
Wald CI	[−0.93, −0.10]	[−1.59, −0.09]	[−1.11, 0.00]	[−1.41, 0.00]	[−0.97, −0.05]	[−1.52, −0.05]
AR CI	[−1.26, −0.21]	[..., −0.31]	[..., −0.14]	[..., −0.18]	[..., −0.18]	[..., −-0.26]
Nonattainment	Imputed	Imputed	Official	Official	Interacted	Interacted
Fixed monitor network	No	Yes	No	Yes	No	Yes

We then replicate all of the key specifications from Chay and Greenstone (2005) and compare prior results of the literature with our new results, correcting for the aforementioned biases. Table 10 compares estimates cross seven specifications from Chay and Greenstone (2005). We explore how our estimates compare with the leading estimates in the literature. The first specification includes no controls. The second specification adds controls for county-level socioeconomic and

[14]This figure was computed by valuing a statistical life at 7 million dollars.

demographic characteristics. Specification 3 includes squared and cubic terms of the prior controls. Specification 4 adds controls for census region fixed effects. Specifications 5 weights observations by county-level population sizes, while specifications 6 controls for covariates in 1970 levels. Finally, specification 7 limits the sample to counties in SMSAs.

Once again, we find that prior estimates were systematically and significantly biased downwards. Across all seven specifications, the leading estimates in the literature significantly understated the impact of air pollution regulation on housing prices, largely due to the failure to account for the non-random placement in monitoring sites. Indeed, looking at the specification in column (5) as an example, we find that a reduction of 1 μg/m3 in air pollution causes, on average, an increase of 0.56% in the median property value (statistically significant at a 5% level), compared with a 0.38% increase in housing prices in Chay and Greenstone (2005) for the same reduction. Thus, prior estimates understate the true benefits of air pollution reduction by nearly 40%. We note that the effect is approximately the same when employing the interacted instrument over the imputed designation. However, this new instrument brings larger uncertainty in the first stage.

Our results show that counties that experienced the largest declines in average TSP levels saw significant increases in median house value over the 10-year period. To put this in perspective, the average median house price across the United States was 30,470 dollars in 1970. Hence, using our results from the fifth specification, a reduction of 10μg/m3 in average TSP levels over the decade would have caused median housing prices to increase by over $1,700 dollars. Converting this result into 2026 dollars, it is equivalent to an increase in median house prices of over $ 14,000 after adjusting for inflation since 1970.

Table 10

	Main specifications				Additional specifications		
Mean TSPs (1/100)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			A. Chay and Greenstone's estimates				
Pt. Estimate	−0.362	−0.213	−0.266	−0.202	−0.379	−0.481	−0.412
Std. Error	(0.152)	(0.096)	(0.104)	(0.090)	(0.186)	(0.183)	(0.269)
1st Stage F-Stat	[31.3]	[29.9]	[24.4]	[21.5]	n/a	n/a	n/a
			B. Replication estimates				
Pt. Estimate	−0.381	−0.281	−0.229	−0.181	−0.398	−0.513	−0.492
Std. Error	(0.171)	(0.123)	(0.095)	(0.078)	(0.158)	(0.212)	(0.191)
AR CI	[−0.85,−0.10]	[−0.62,−0.08]	[−0.47,−0.07]	[−0.38,−0.04]	[−0.89,−0.15]	[−1.26,−0.21]	[−1.07,−0.20]
1st Stage F-Stat	[24.51]	[21.88]	[20.69]	[17.51]	[12.49]	[13.88]	[23.11]
			C. Fixed monitor network				
Pt. Estimate	−0.679	−0.452	−0.365	−0.272	−0.520	−0.841	−0.744
Std. Error	(0.346)	(0.206)	(0.159)	(0.120)	(0.188)	(0.383)	(0.314)
AR CI	[−1.95,−0.17]	[−1.10,−0.12]	[−0.82,−0.11]	[−0.6,−0.07]	[−1.11,−0.22]	[. . . ,−0.31]	[−1.95,−0.28]
1st Stage F-Stat	[12.94]	[15.60]	[13.06]	[13.99]	[10.94]	[10.09]	[11.69]
			D. [Fixed monitor network with interacted IV]				
Pt. Estimate	−0.583	−0.458	−0.328	−0.208	−0.563	−0.781	−0.939
Std. Error	(0.324)	(0.222)	(0.176)	(0.132)	(0.226)	(0.378)	(0.446)
AR CI	[−1.72,−0.11]	[−1.19,−0.11]	[−0.91,−0.04]	[−0.61,−0.02]	[. . .,−0.21]	[. . .,−0.26]	[. . .,−0.36]
1st Stage F-Stat	[13.43]	[14.54]	[9.87]	[10.27]	[7.82]	[9.61]	[7.19]
County Data Book Covariates		✓	✓	✓	✓		✓
Flexible Forms for Covariates			✓	✓	✓		
Census Region Fixed Effects				✓	✓		
Covariates in 1970 Levels						✓	
Population Weighting					✓		
Counties in SMSAs Only							✓
Sample Size (# Counties)	988	980	980	980	980	981	385

Notes: Standard errors in parentheses. Anderson-Rubin 95% confidence sets in brackets. First-stage F-statistics in square brackets.

8 Other Implications For the Literature

We investigate some other implications in this section.

As noted, in a two-stage least squares setup, the consequence of an inflated first stage is an attenuated second stage. By overstating the effects of nonattainment on TSP reductions, we have seen how the literature estimates effects of regulation-induced reductions in TSP on health and housing values that are biased down. Here, we discuss implications of this first-stage finding for other substantive contributions of the literature, beginning with several relatively recent studies.

Long-run labor market earnings. Isen, Rossin-Slater, and Walker (2015) use cross-county variation in regulatory effort induced by the 1970 Amendments to measure the effects of *in utero* and early childhoold exposure to TSP on labor market earnings later in life. They find that a 10 $\mu g/m^3$ reduction in annual average exposure to TSP while young raises labor income at age 29-31 by around 1 percent. These effects are large when compared to the total monetized benefits of

reducing air pollution. Doubling them could thus rationalize much tighter air quality regulation than might otherwise be adopted.

Fetal mortality. Sanders and Stoecker (2015) estimate the effects of exposure to air pollution on fetal mortality. As no reliable fetal death statistics exist, they infer fetal deaths from changes in infant sex ratios. Male fetuses tend to be more vulnerable to health shocks than female fetuses, so an increase in the share of live births that are male would suggest a reduction in fetal mortality, all else equal. Sanders and Stoecker (2015) estimate that a 1 $\mu g/m^3$, regulation-induced reduction in a county's TSP concentration increases the probability that a live birth is male by 0.029 to 0.043 of a percentage point. It is unclear whether doubling this estimate would significantly affect cost-benefit calculations of alternative air quality standards (for example), as the evidence base necessary to place a dollar value on fetal deaths has not been developed.[15]

Infant mortality. Using the same first-difference specification focusing on changes between 1971 and 1972, Chay and Greenstone (2003a) find that a 1 percent decline in a county's annual average TSP concentration leads to a 0.5 percent decline in its infant mortality rate. The corresponding elasticity estimated by Chay and Greenstone (2003b), who exploit sharp changes in TSP concentrations caused by the 1981-1982 recession, is that a 1 percent decline in a county's annual average TSP concentration leads to a 0.35 percent decline in its infant mortality rate – opposite the expected pattern, which is that estimates from the Clean Air Act would be half those from recession. It is, however, difficult to draw firm conclusions because estimates from the Clean Air Act are imprecise. In particular, it is not possible to reject that estimates from the Clean Air Act are half those from the recession. The most recent studies using the 1970 Amendments as a research design bring in additional years of post-reform data to increase precision.

9 Conclusion

Pollution levels in several developing countries now equal or exceed those experienced in the United States in 1970. In their efforts to balance environmental quality and economic growth, regulators in those countries will require knowledge of the benefits available from reducing air pollution in a high-pollution context. Given the stakes involved in setting regulatory targets effectively in such contexts, our analysis offers a new perspective on the widely-used variation in regulations due to the Clean Air Act Amendments of 1970 and the benefits of particulate pollution reductions that these generated.

Our analysis has made three contributions. First, drawing on data thought no longer to exist, we showed how attainment status under the act is defined at the Air Quality Control Region,

[15] As Sanders and Stoecker (2015) state in footnote 27: "While miscarriages and fetal deaths carry with them both medical and substantial psychological costs, some may occur without the knowledge of the mother. Of the fetal deaths that occur in known pregnancies, the argument could be made that actual financial costs should be less than that of an infant death, as less time and fewer resources have been invested, but it remains difficult to quantify the costs associated with the potentially substantial psychological impacts."

not the county. This leads us to a new designation of regulatory treatment under the act, one that has greater power when examining the effects of the legislation on pollution levels. Related to treatment designation, we argue that institutional details do not support use of a regression-discontinuity estimator in this setting.

Second, we have shown how changes in the monitoring network need to be account for when analysing the effects of the legislation. Estimates from a popular difference-in-differences estimator of the impact of the regulations on pollution will tend to be biased due to the growth and churning in the TSP monitor network around the 1970 Amendments, associated with regulatory 'gaming' behavior. Further, we show how accurate estimates of the effects of regulation on air pollution can be obtained by using a balanced panel of TSP monitors to define county-level air pollution trends.

Third, we presented new estimates of the causal impacts of air pollution based on the new measure of treatment and accounting for systematic changes in the monitoring network.Taking the new first-stage estimates based on the balanced monitor panel and additional regulatory variation, we showed how 2SLS estimates of both health outcomes based on a long-difference design and house values exceeded well-known estimates in the literature by a significant margin. In turn, the larger revised estimates we obtain can rationalize much more stringent air quality regulation than prior estimates using this source of regulatory variation suggested.

References

Chay, K., C. Dobkin, and M. Greenstone (2003). The Clean Air Act of 1970 and Adult Mortality. *The Journal of Risk and Uncertainty 27*(3), 279–300.

Chay, K. and M. Greenstone (2003). The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *Quarterly Journal of Economics 118*(3), 1121–1167.

Chay, K. and M. Greenstone (2005). Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy 113*(2), 376–424.

Graff Zivin, J. and M. Neidell (2013). Environment, Health, and Human Capital. *Journal of Economic Literature 51*(3), 689–730.

Isen, A., M. Rossin-Slater, and W. R. Walker (2017). Every Breath You Take, Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970. *Journal of Public Economics 125*(3), 898–902.

Sanders, N. and C. Stoecker (2015). Where have all the young men gone? Using fetal sex ratios to measure fetal death rates. *Journal of Health Economics 41*, 30 – 45.

Smith, V. K. (2007). Reflections on the literature. *Review of Environmental Economics and Policy 1*(2), 300–318.

Supplemental Material

This Supplemental Material consists of the following sections: Section **??** provides ... And Section A presents additional tables and figures.

A Additional Tables and Figures

Table A1

	Coefficient on indicator for official nontattainment in 1972					
	Fixed set of sites			Time varying set of sites		
Dependent variable	No weights (1)	Site observation weights (2)	County population weights (3)	No weights (4)	Site observation weights (5)	County population weights (6)
Change 1971-1972	-3.027	-2.663	-4.678	-2.672	-3.097	-6.742
n=488	(1.495)	(1.450)	(2.304)	(2.435)	(2.256)	(3.497)
	[4.1]	[3.371]	[4.123]	[1.205]	[1.885]	[3.718]
Change 1971-1973	-3.359	-3.237	-3.507	-3.552	-3.142	-6.524
n=477	(2.152)	(2.177)	(3.846)	(2.626)	(2.360)	(4.897)
	[2.436]	[2.211]	[.831]	[1.83]	[1.772]	[1.774]
Change 1971-1974	-3.609	-3.659	-9.683	-3.916	-3.891	-6.589
n=462	(1.994)	(1.990)	(3.517)	(2.908)	(2.674)	(5.266)
	[3.276]	[3.381]	[7.579]	[1.814]	[2.118]	[1.566]

Notes: Notes: F-statistic corresponding critical value 3.863 Each estimate derives from a separate regression. Heteroskedasticity-robust standard errors are given in parentheses, and F-statistics in square brackets. Sample size, referring to the number of counties with at least one monitor active in both years, is reported in the row heading. Inputted attainment status designations vary at the AQCR region.

Table A2

	Coefficient on indicator for interacted nontattainment in 1972					
	Fixed set of sites			Time varying set of sites		
Dependent variable	No weights (1)	Site observation weights (2)	County population weights (3)	No weights (4)	Site observation weights (5)	County population weights (6)
Change 1971-1972	-3.430	-3.325	-4.312	-7.989	-7.861	-9.889
n=488	(1.362)	(1.296)	(1.953)	(2.288)	(2.086)	(2.550)
	[6.342]	[6.58]	[4.875]	[12.191]	[14.203]	[15.043]
Change 1971-1973	-5.878	-6.068	-6.553	-10.458	-10.154	-11.913
n=477	(1.570)	(1.570)	(2.813)	(2.420)	(2.236)	(2.924)
	[14.015]	[14.946]	[5.428]	[18.675]	[20.629]	[16.602]
Change 1971-1974	-4.967	-5.351	-9.065	-10.857	-11.086	-13.531
n=462	(1.628)	(1.616)	(2.750)	(2.525)	(2.362)	(3.218)
	[9.309]	[10.962]	[10.865]	[18.491]	[22.035]	[17.678]

Notes: Notes: F-statistic corresponding critical value 3.863 Each estimate derives from a separate regression. Heteroskedasticity-robust standard errors are given in parentheses, and F-statistics in square brackets. Sample size, referring to the number of counties with at least one monitor active in both years, is reported in the row heading. Interacted attainment status is imputed * official status.