Online Appendix
Uncharted Waters: Effects of Maritime Emission Regulation

Jamie Hansen-Lewis and Michelle Marcus

Contents

A Data

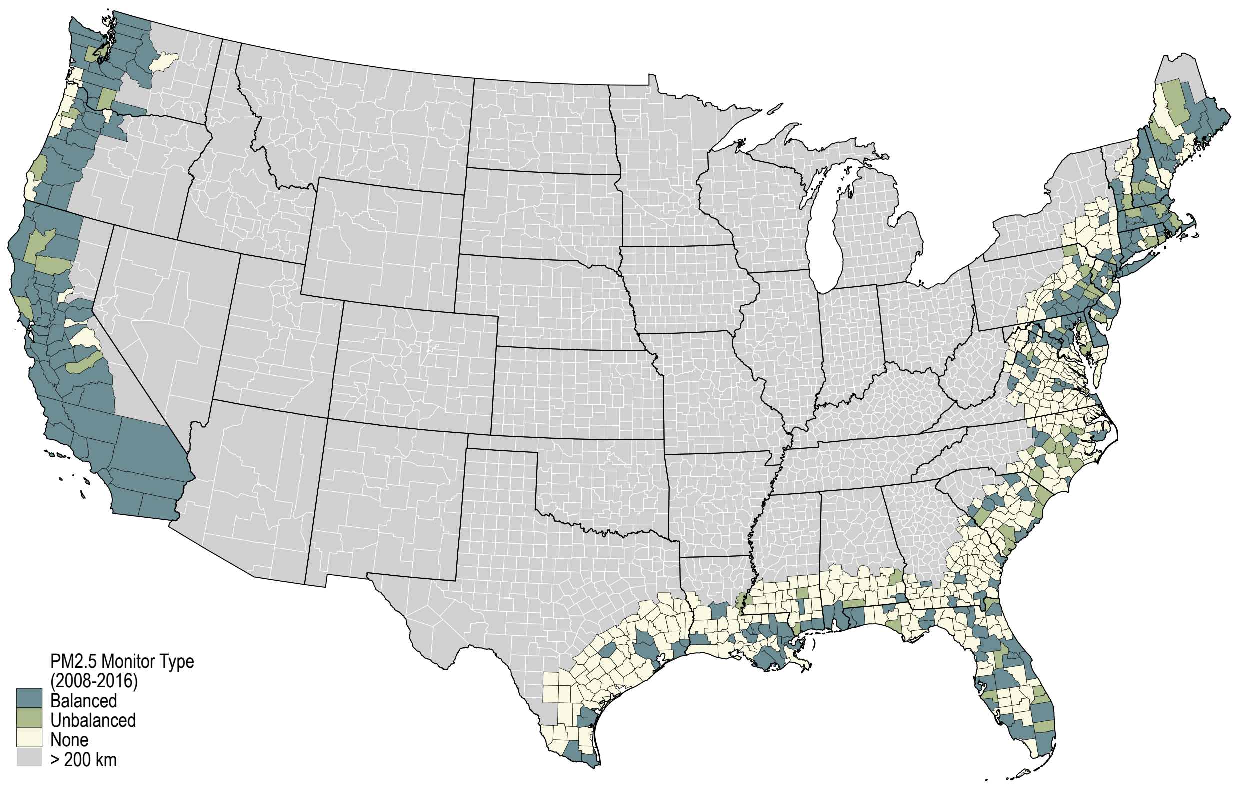
A.1 Air Pollution Data

The source is the US EPA Air Quality System (AQS). The AQS records provide daily summaries from outdoor air quality monitors across the United States for a variety of pollutants.[24] Raw observations are at the level of the pollutant-monitor-day. We construct PM2.5-monitor-day observations from 3 PM2.5 from pollutant codes. Our primary source is PM2.5 coded as pollutant 88101. For monitor-days where 88101 data are missing we substitute with PM2.5 coded as pollutant 88502. When both 88101 and 88502 are missing, we substitute with 88501. Thus, we obtain PM2.5-monitor-day observations.

[24] AQS data are collected to ensure compliance with state and federal air quality regulations as well as to support air pollution research. They are the principal source of historical air quality and have been previously employed in numerous studies (Fann, Wesson and Hubbell, 2016).

Monitor-day observations are collapsed to monitor-week averages. The monitors are matched to the county in which the monitor is located. To construct a balanced panel of monitors, monitors that are not observed for at least one week each year from 2008-2016 are dropped. We then average the remaining monitor-weeks within each county to construct county-week observations of mean PM2.5. Last, we collapse county-week observations to county-month averages. Throughout, averages exclude missing observations. Figure A1 depicts the 232 counties in the balanced panel of monitors and the counties with a monitor not in the balanced panel.

Figure A1: Analysis Sample



Note: Figure shows non-grey counties with population-weighted centroids within 200km of heavy ship traffic, as defined by the top 5th percentile of 2011 vessel density raster grid cells. Counties with population-weighted centroids further than 200km are shaded in grey. Blue counties are those with a balanced PM2.5 monitor. They have at least one PM2.5 monitor with at least one observation per year from 2008 to 2016. Green counties are those with only unbalanced PM2.5 monitors. They have PM2.5 monitor(s) but no single monitor with at least one observation per year from 2008 to 2016. Yellow counties have no PM2.5 monitors.

A.2 NAAQS Standards

We use two data sets on counties' air quality performance relative to the National Ambient Air Quality Standards (NAAQS). In robustness checks, we control for county attainment status. We obtain attainment status for each pollutant, standard, county, and calendar year from the US EPA Green Book (U.S. EPA, 2023*c*). We focus explicitly on PM2.5 1997, 2006, and 2012 standards; PM10 1987 standards; sulfur dioxide 1971 and 2010 standards; nitrogen dioxide 1971 standards; ozone 1979, 1997, 2008, and 2015 standards; and carbon monoxide 1971 standards. We include as controls indicators for whether part or all of the county is in non-attainment of any of the listed standards for each pollutant. In our analysis of behavioral responses, we classify counties based on their degree of compliance or non-compliance with the NAAQS PM2.5 standards in 2012. To determine compliance with the NAAQS, the US EPA requires raw monitoring data to meet stringent quality standards and follows particular formulas for aggregating. We employ the EPA's output of these calculations, called the design values. We obtain the cross-section of the 2012 PM2.5 design values, based on data from 2010-2012, for each country and standard (24-hour and annual) from the US EPA (U.S. EPA, 2022*a*).

A.3 Weather Data

We use the PRISM Daily Weather Data for the Contiguous United States (Schlenker, 2020). We compute the county-day means for each weather variable as the average of the grid-cell-day observations within the county. We calculate cubic functions of county-day minimum temperature, maximum temperature, and total precipitation, as well as the interactions of precipitation with minimum temperature and maximum temperature. Last, we average over the county-day observations to form county-month observations for each weather variable for our baseline weather controls.

A.4 Outdoor Activity Data

We use two sources of data to measure outdoor activity in order to observe whether individuals exhibit behavioral changes in response to changing air quality. First, we make use of recreation data from Recreation.gov, which maintains data on millions of visitors to federal parks. We use data on campsite reservations from 2008 to 2016, which include over 24 million individual reservations at over 3,400 facilities. We limit the sample to campsites in the continental US.[25] We collapse the visit-level data to the facility-by-month level and focus on number of visits, total people visiting, and number of days.

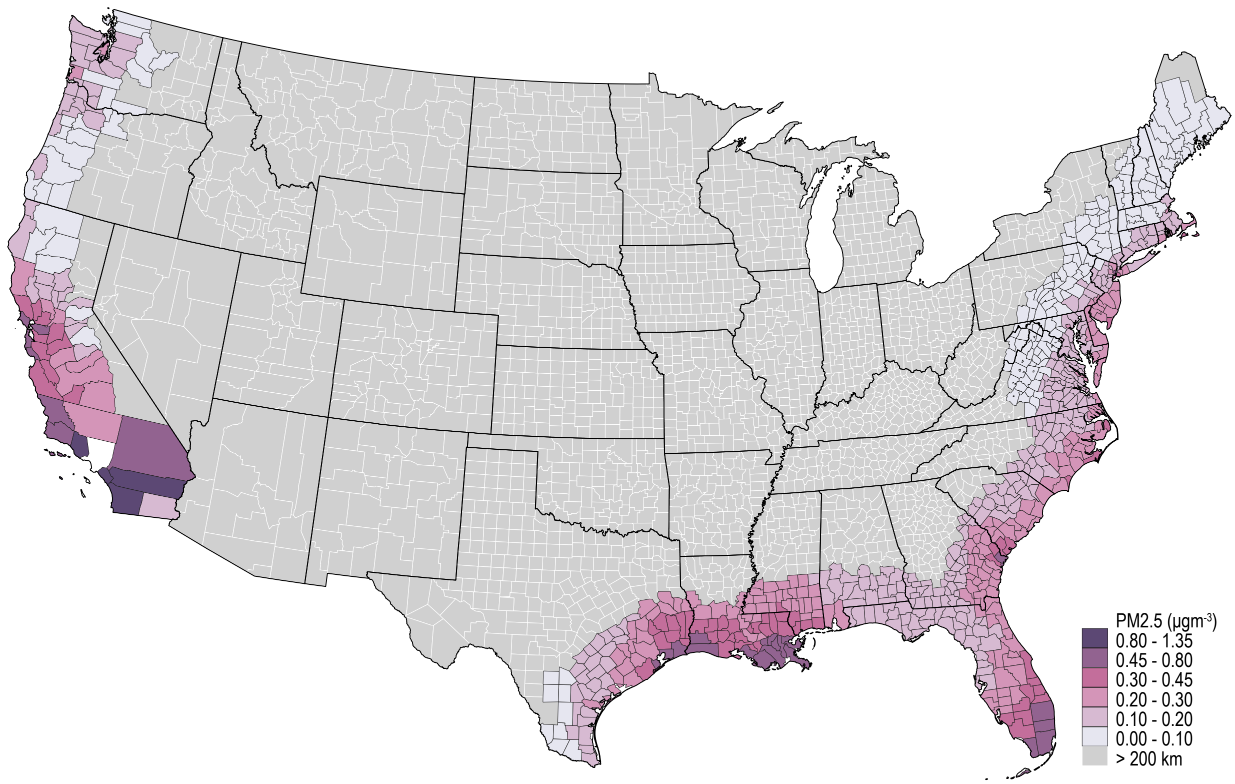
We supplement this with data from the American Time Use Survey (ATUS) from 2008 to 2016. Conducted by the US Census Bureau and the Bureau of Labor Statistics, the ATUS asks respondents to provide a detailed time diary of all activities over a 24-hour period, including the location of each activity. We use the location information to measure respondents' time spent outdoors. Additional information records respondents' county of residence, gender, race, ethnicity, education, age, presence of a child in the household, and information on the day of the week and whether the survey was conducted on a holiday.

[25] About 94 percent of facilities are classified as "sites." The remaining categories include facilities classified as entrance, lottery, POS, and tour. We exclude these categories to capture a homogeneous set of campsites where we are confident that visitors are spending time outdoors, but the results are robust to including the other categories.

B ECA Effect Additional Results and Robustness

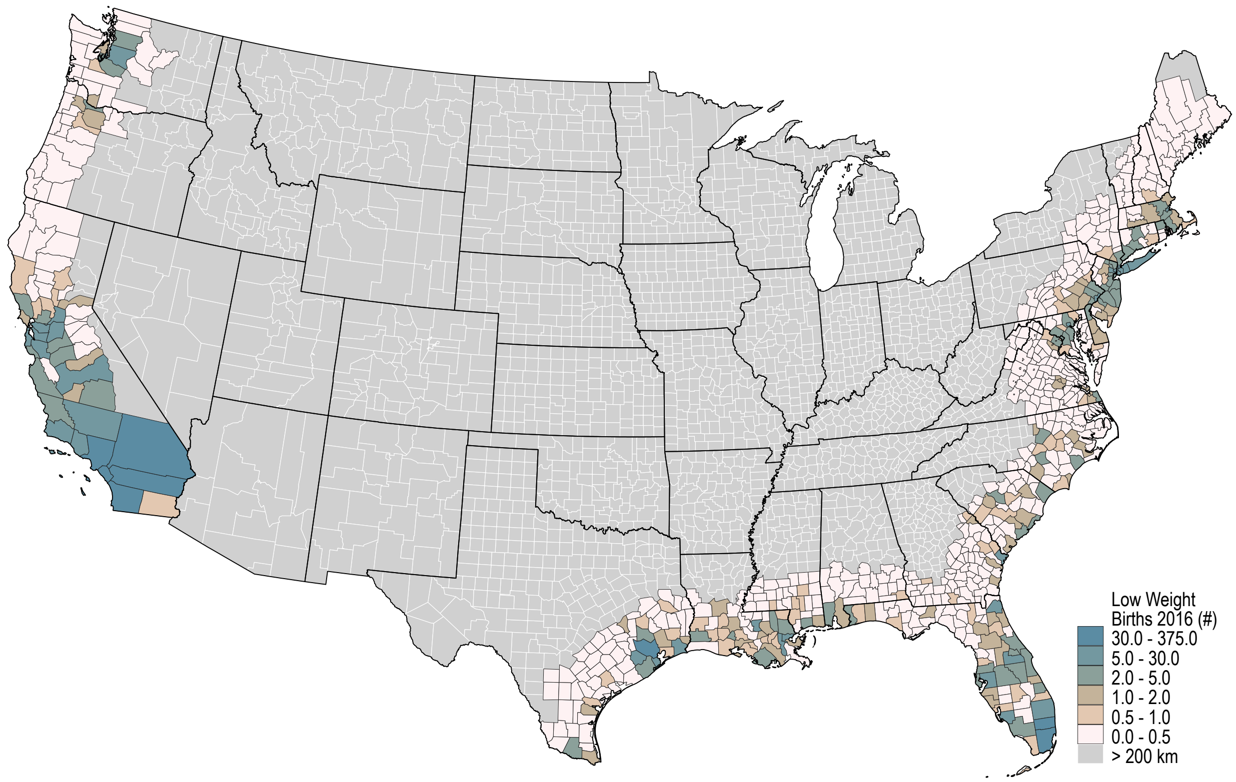
B.1 Spatial Distribution of Main Effects

Figure A2: Scaled Reduction in PM2.5



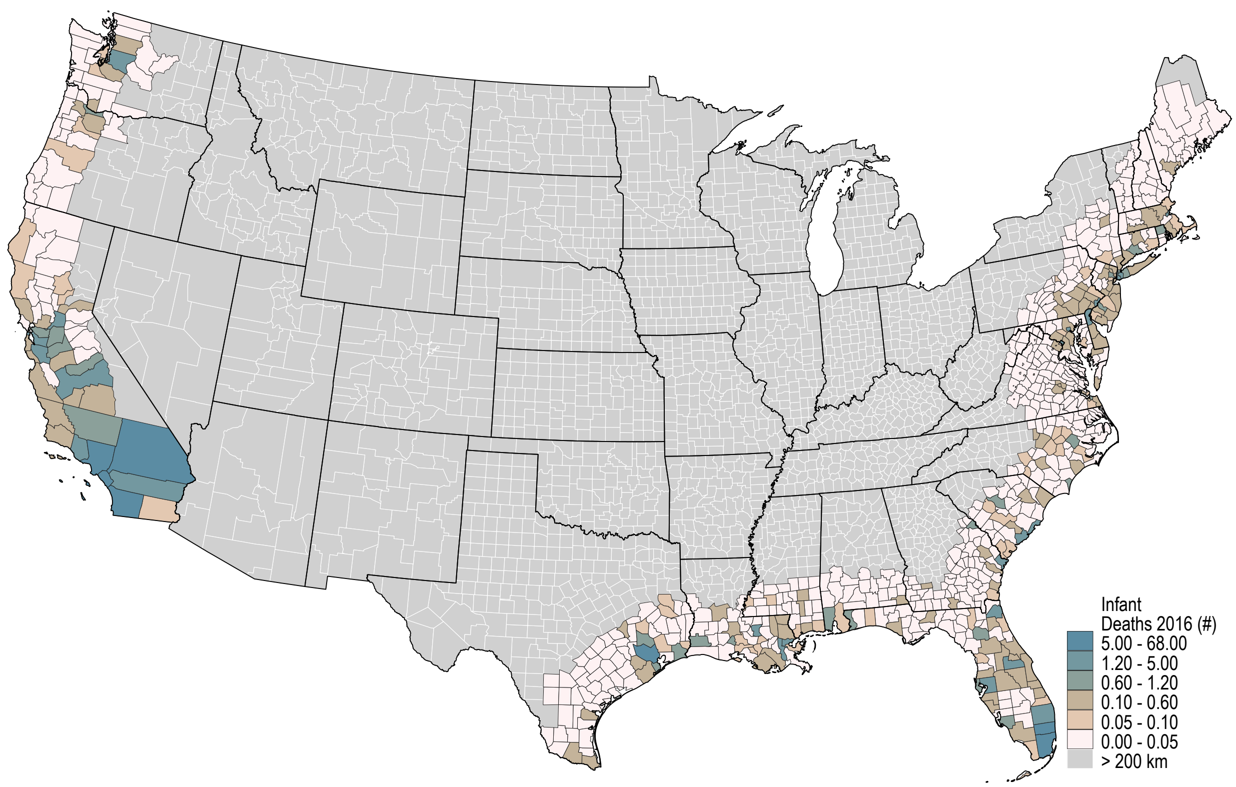
Note: Figure shows the estimated reduction in ambient PM2.5 from the ECA at the county level. The estimated reductions are the county level CMAQ predictions depicted in Figure 3 scaled by the estimated ECA effect coefficient in Table 2 column 1.

Figure A3: Reductions in Low Birth Weight Infants



Note: Figure shows the estimated reduction in low birth weight (<2,500 g) infants at the county-level from the ECA policy in 2016. See text for details.

Figure A4: Reductions in Infant Deaths



Note: Figure shows the estimated reduction in infant deaths at the county-level from the ECA policy in 2016. See text for details.

B.2 Additional Health Outcomes

We first examine other indicators of infant health and adult health. Columns 1-2 of Table A1 show significant improvements from the ECA on average birth weight and gestation. As the elderly also tend to be particularly sensitive to air pollution, we explore the effects on elderly mortality. Columns 3-4 of Table A1 and Figure A5 show that the policy led to statistically significant declines in mortality for individuals age 75-84 and age 85 and over. A one-unit predicted change in PM2.5 from CMAQ led to declines in elderly mortality of 0.03 and 0.15 percentage points, or 0.8 and 1.4 percent for ages 75-84 and above 85, respectively. Panels (a) and (b) of Figure A5 show little evidence of pre-trends in years prior to the policy for elderly mortality and a decrease in mortality in areas with heavy ship traffic after the ECA policy.

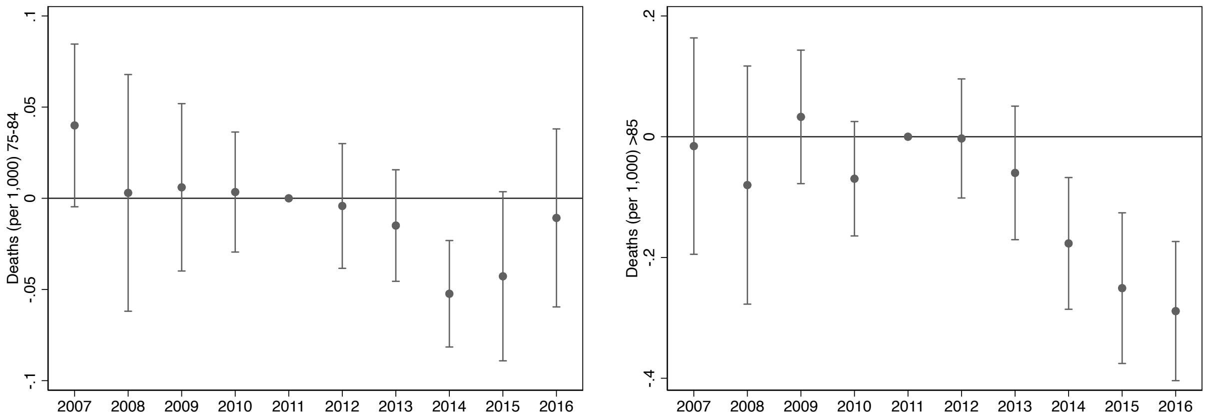
Similarly, We explore the distributional effect on birth weight further in the reduced form results in Panel A of Table A2, which shows the effect of the policy on bins of birth weight. Consistent with the stronger effects on infant health at the lower end of the distribution, we find large reductions in births for the four smallest bins in the birth weight distribution and increases in births in the middle of the distribution. These results suggest that there are important impacts not only for low birth weight (less than 2,500 g) infants, but also very low birth weight (less than 1,500 g) and extremely low birth weight (less than 1,000 g) infants. The negative health consequences are especially severe for very and extremely low birth weight infants, so improvements in these categories are quite beneficial.

Table A1: **Effects of ECA on Additional Health Outcomes**

	(1) Birth weight	(2) Gestation	(3) Deaths: 75-84	(4) Deaths: >85
Panel A. Reduced Form				
Post-ECA*CMAQ	1.620	0.012	-0.031	-0.151
	(0.874)*	(0.006)**	(0.010)***	(0.043)***
R^2	0.82	0.71	0.77	0.65
N	25,052	25,052	25,056	25,056
N-counties	232	232	232	232
Mean	3305.18	38.78	3.72	10.90
%Change	0.05	0.03	-0.83	-1.38
Panel B. 2SLS				
PM2.5	-3.309	-0.024	0.056	0.278
	(1.835)*	(0.011)**	(0.019)***	(0.078)***
R^2	0.81	0.67	0.76	0.60
N	24,901	24,901	24,905	24,905
F	18.33	18.33	27.54	26.91
N-counties	232	232	232	232
Mean	3305.10	38.78	3.72	10.90
%Change	-0.10	-0.06	1.50	2.55
Panel B. OLS				
PM2.5	0.087	-0.000	0.009	0.033
	(0.079)	(0.000)	(0.002)***	(0.004)***
R^2	0.82	0.71	0.77	0.65
N	24,901	24,901	24,905	24,905
N-counties	232	232	232	232
Mean	3305.10	38.78	3.72	10.90
%Change Post-ECA	0.00	-0.00	0.24	0.31

Note: Columns 1 and 2 repeat the analysis of Table 3 column 1 with outcomes birth weight in grams (column 1) and gestation in weeks (column 2). Columns 3 and 4 repeat the analysis of Table 3 column 4. In column 3, the outcome is deaths per 1,000 among individuals aged 75 to 84 and the observations are weighted by the population aged 75 to 84. In column 4, the outcome is deaths per 1,000 among individuals aged 85 and older and the observations are weighted by the population aged 85 and older.

Figure A5: **Effects of ECA on Elderly Mortality**



(a) 75-84

(b) 85 and over

Note: Repeats the analysis of Figure 6. In column 1, the outcome is deaths per 1,000 among individuals aged 75 to 84 and the observations are weighted by the population aged 75 to 84. In column 2, the outcome is deaths per 1,000 among individuals aged 85 and older and the observations are weighted by the population aged 85 and older.

Table A2: **Effects of ECA on Birth Weight Distribution**

	(1) <1,000 g	(2) 1,000-1,500 g	(3) 1,500-2,000 g	(4) 2,000-2,500 g	(5) 2,500-3,000 g	(6) 3,000-3,500 g	(7) 3,500-4,000 g	(8) 4,000-4,500 g	(9) >4,500 g
Panel A. Reduced Form									
Post-ECA*CMAQ	-0.173	-0.111	-0.266	-0.776	-0.347	1.807	0.456	-0.466	-0.124
	(0.075)**	(0.073)	(0.115)**	(0.201)***	(0.593)	(0.483)***	(0.472)	(0.292)	(0.141)
R^2	0.26	0.16	0.20	0.42	0.61	0.34	0.60	0.58	0.28
N	25,052	25,052	25,052	25,052	25,052	25,052	25,052	25,052	25,052
N-counties	232	232	232	232	232	232	232	232	232
Mean	5.28	5.29	10.62	39.34	178.41	403.91	276.41	69.85	10.89
%Change	-3.27	-2.10	-2.50	-1.97	-0.19	0.45	0.16	-0.67	-1.14
Panel B. 2SLS									
PM2.5	0.350	0.225	0.534	1.566	0.746	-3.647	-0.930	0.908	0.248
	(0.169)**	(0.170)	(0.270)**	(0.620)**	(1.112)	(1.209)***	(0.959)	(0.637)	(0.304)
R^2	0.22	0.14	0.15	0.34	0.61	0.26	0.60	0.57	0.27
N	24,901	24,901	24,901	24,901	24,901	24,901	24,901	24,901	24,901
F	18.33	18.33	18.33	18.33	18.33	18.33	18.33	18.33	18.33
N-counties	232	232	232	232	232	232	232	232	232
Mean	5.28	5.29	10.62	39.35	178.44	403.94	276.38	69.82	10.89
%Change	6.62	4.25	5.02	3.98	0.42	-0.90	-0.34	1.30	2.28
Panel C. OLS									
PM2.5	-0.007	-0.004	-0.006	0.012	-0.026	-0.080	0.129	-0.026	0.007
	(0.010)	(0.010)	(0.015)	(0.028)	(0.056)	(0.056)	(0.058)**	(0.034)	(0.016)
R^2	0.26	0.16	0.20	0.42	0.61	0.34	0.60	0.58	0.28
N	24,901	24,901	24,901	24,901	24,901	24,901	24,901	24,901	24,901
N-counties	232	232	232	232	232	232	232	232	232
Mean	5.28	5.29	10.62	39.35	178.44	403.94	276.38	69.82	10.89
%Change Post-ECA	-0.13	-0.07	-0.05	0.03	-0.01	-0.02	0.05	-0.04	0.07

Note: Repeats the analysis of Table 3 column 1 with outcomes of births per 1,000 in 500 gram intervals of the domain of birth weight.

B.3 Other Pollutants

In Table A3, we provide estimates of the impact of the policy on other pollutants and the AQI using data from U.S. EPA (2022*c*) and U.S. EPA (2022*d*). We expect the spatial and temporal pattern of these pollutants' impacts will be different than those of PM2.5, and each other, because each pollutant has a distinct chemistry that determines how it is dispersed, deposited, or converted to other pollutants. While the CMAQ model predictions reflect the expected atmospheric dispersion for PM2.5 from the policy, this may not accurately capture the spatial impacts of other pollutants. Given the relatively short atmospheric lifetime of SO2 before it reacts in the atmosphere, we do not expect reductions in this pollutant to be as widely geographically distributed as the reductions in PM2.5.[26] Consistent with this prior, we fail to see an impact of the ECA on SO2 using the predicted PM2.5 decline from the CMAQ model as our measure of intensity of exposure to the policy in column 1, but we find more precise evidence suggestive of a modest SO2 improvement when using distance as the definition of exposure (column 2). Next, column 3 shows there is a significant decline in NO2. While the ECA's engine requirements targeting NOX likely contributed to this decline, we expect the ECA's contribution to have been small because the engine requirements would have phased in for well under 25% of the US fleet during the sample (p.2-40). As an additional check to isolate the effect of the ECA, row 12 of Table A5 shows that the reduced form effects of the ECA on PM2.5, low infant birth weight, and infant death are robust to including NO2 as a control variable.

In addition to the primary pollutants NO2 and SO2, we also look for effects of the ECA on ozone (O3) because it was a secondary pollutant of interest to the regulator. In column 4, we report that we failed to see an impact of the ECA on ozone using the predicted PM2.5 decline from the CMAQ model as our measure of intensity of exposure.[27] The null result in column 4 indicates that any ozone effect is not so strongly correlated with the PM2.5 effect as to explain our health results.

[26] SO2 rapidly dissolves in water droplets in the air or reacts to form sulfuric acid gas. The rate of these processes depends on atmospheric conditions. As an example, in a cloud, 60% of SO2 gas molecules are converted to other molecules within 20 minutes (Jacobson, 2002). The atmospheric lifetime of remaining SO2 is approximately 7.2 days (U.S. EPA, 2017).

[27] The ECA may still have yielded a small improvement in ozone, but any effect would be small. An alternative research design for this outcome would be to focus on seasonal ozone and use the separate CMAQ output for this pollutant rather than PM2.5. Still, we expect these effects to be small because the regulator's analysis predicted at most a 1% decline in summer season 8-hour max ozone (p. 3-28).

Table A3: **Effect of ECA on Other Pollutants**

	(1) SO2 (AQI)	(2) SO2 (AQI)	(3) NO2 (AQI)	(4) O3 (AQI)	(5) AQI
Post-ECA*CMAQ	-0.200		-0.933	0.390	-1.270
	(0.357)		(0.272)***	(0.366)	(0.416)***
Post-ECA*Dist		-0.014			
		(0.005)***			
R^2	0.75	0.75	0.89	0.85	0.88
N	7,273	7,273	8,372	17,722	24,991
N-counties	68	68	79	184	232
Mean	3.61	3.61	22.15	39.20	54.32
%Change	-5.54	-0.39	-4.21	1.00	-2.34

Note: The unit of observation is county-year-month. The observations are weighted by the number of births conceived in county i in year-month ym. The sample includes counties with population-weighted centroids within 200km of heavy ship traffic and with a PM2.5 monitor from 2008-2016. In columns 1-4 the sample is further restricted to counties with at least one monitor for the corresponding outcome pollutant from 2008-2016. Reduced-form estimates are obtained from equation 1 with outcomes of monthly mean sulfur dioxide (column 1), nitrogen dioxide (column 3), ozone (column 4), and maximum AQI across all criteria pollutants and standards (column 5). Column 2 repeats column 1 with the the intensity of exposure to the policy is measured by distance to the nearest major port in lieu of CMAQ prediction in equation 1. Air quality index maps physical pollutant concentrations to a 0–500 scale according to the health risk (U.S. EPA, 2018). Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

B.4 Placebo Outcomes

Next, we examine a number of placebo outcomes that support the validity of our results. A potential concern with our estimation of the ECA's effects on health at birth is that the introduction of the ECA could have been correlated with changes in demographic characteristics or local economic activity. For example, if the introduction of the ECA was correlated with an increase in conceptions for mothers with high proclivity for prenatal care in coastal counties, then our results reflect the change in the composition of mothers rather than the change from the air quality improvement the policy induced.

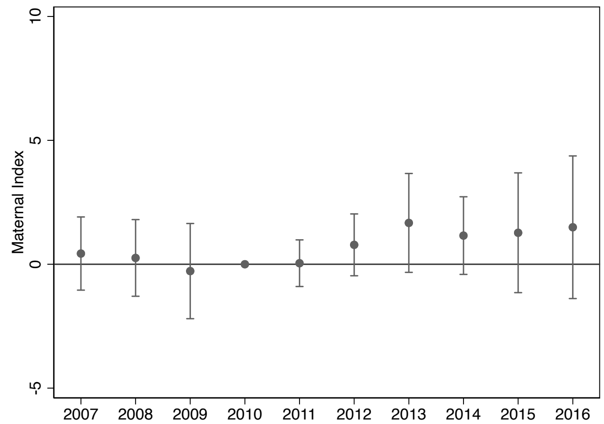
We failed to find evidence that demographic and economic shocks correlated with the treatment are driving the results in Table 2. To represent the many demographic characteristics relevant to infant health, we construct an index of maternal demographic characteristics, defined as the birth weight predicted from only observed maternal characteristics, including education, marital status, race, ethnicity, age, smoking status, and diabetes. As shown in Table 2 column (2) and Figure A6, we failed to find that these demographic characteristics are changing simultaneously with policy exposure. Similarly, column (3) shows that the policy did not result in a significant change in the number of conceptions. Finally, column (4) shows there is no evidence that the policy is correlated with differential changes in economic activity as measured by the unemployment rate. In addition, Table A4 shows there is no significant relationship between the policy variation and the following additional outcomes: log pollutant emissions from power plants (Clean Air Markets Program Data from U.S. EPA (2022*b*)), log emissions of PM2.5 (National Emissions Inventory from U.S. EPA (2023*a*)), the number of other Toxic Release Inventory (TRI) pollution sources (County Business Patterns Data from U.S. Census Bureau (2023)), the frequency of monitor readings, employment, and earnings. These results provide additional support in favor of the assumption that the policy instrument captures only changes in pollution, rather than changes in other confounding drivers of health.

Table A4: **Effects of ECA on Placebo Outcomes**

	(1) log(elec. emissions)	(2) log(PM2.5 emissions)	(3) TRI Establishments (per 1,000)	(4) N $PM_{2.5}$ obs	(5) Employees (per 1,000)	(6) Payroll ($1,000 per 1,000)
Post-ECA*CMAQ	-0.055	-0.024	-0.007	-0.033	0.250	-643.017
	(0.044)	(0.019)	(0.005)	(0.160)	(1.930)	(497.209)
R^2	0.84	0.95	0.99	0.75	0.99	0.98
N	13,739	8,352	25,052	25,052	25,052	25,052
N-counties	140	232	232	232	232	232
Mean	5.09	5.85	0.24	7.03	386.01	20476.12
%Change	-5.47	-2.36	-3.03	-0.47	0.06	-3.14

Note: The unit of observation is the county-year-month. The sample includes counties with population-weighted centroids within 200km of heavy ship traffic and with a PM2.5 monitor from 2008-2016. The observations are weighted by the number of births conceived in county i in year-month ym. Reduced-form estimates are obtained from equation 1. Column 1 reports the effect of the ECA on emissions defined as log of total tons of CO_2, SO_2, and NO_x from power plants reported in the EPA's Clean Air Markets Program data. Column 2 repeats column 1 with the log of PM2.5 emissions from the National Emissions Inventory (NEI) data as the outcome variable. Column 3 repeats column 1 with the number of TRI establishments per 1,000 population from the County Business Patterns (CBP) data as the outcome variable. Column 4 repeats column 1 with the mean monitor-days per week with an observation. Columns 5 and 6 repeat column 1 with the number of employees per 1,000 population and the payroll in thousands of dollars per 1,000 population at any CBP establishment. The insignificant coefficients indicate there is no evidence of changes in underlying economic and pollution characteristics that are correlated with the CMAQ policy variation. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Figure A6: **Maternal Demographics and CMAQ Exposure**



Note: The unit of observation is a county-year-month. The observations are weighted by the number of births conceived in county i in year-month ym. The sample includes counties with population-weighted centroids within 200km of heavy ship traffic and with a PM2.5 monitor. The outcome is predicted birth weight based on observed characteristics and coefficients obtained from regressing birth weight on maternal characteristics, including education, marital status, race, ethnicity, age, smoking status, and diabetes. The depicted coefficients are the estimated effect of a one-unit increase in a county's CMAQ predicted reduction from the ECA in each year relative to the year before the ECA came into effect. Robust standard errors are clustered at the county level. The confidence intervals are $\pm$ 1.96 standard errors.

B.5 Comparison with Previous Approaches

Employing the CMAQ output as a measure of intensity of exposure to the ECA policy improves on approaches that rely on imprecise proxies for source-specific exposure. To illustrate the distinction, we perform our analysis using distance to a port as the proxy for ship pollution exposure in lieu of CMAQ output. We define distance to a port as the kilometers from the county population-weighted centroid to the nearest major US port.[28] We highlight two main concerns with distance metrics. First, distance is a poor proxy for exposure to improvements from the policy because atmospheric interactions play a major role in the dispersion of pollution. This concern would lead to bias from measurement error. Second, there does not exist an *a priori* functional form for the relationship between the distance from a pollution source and pollutant exposure from the source. This concern would lead to bias from misspecification.

Table A7 reports the results of estimating equation (1) where intensity of exposure to the policy is measured by either CMAQ or distance. We report the results for infant deaths, low birth weight, and fine particulate matter in panels A-C, respectively. We standardize the coefficients and standard errors into units of standard deviations so that the results are comparable across candidate treatment variables. First, we compare the Bayesian information criteria (BIC), which is a criterion for model selection based, in part, on the likelihood function. We observe that the estimates based on CMAQ consistently yield a lower BIC across all outcomes, suggesting that CMAQ is preferred. Across all outcomes, the CMAQ model appears to reduce measurement error, as expected. The T-statistic is larger and standard errors are smaller for CMAQ relative to distance in all panels. The estimated effect of a one standard deviation increase in distance relative to CMAQ exposure led to a slightly larger reduction in infant deaths and low birth weight, but a slightly smaller reduction in fine particulate matter. However, we do not emphasize these differences because the confidence intervals of these estimates overlap. Nevertheless, these models show meaningful improvements in precision when CMAQ is used to measure exposure to the ECA policy.

B.6 Robustness checks

Table A5 shows our results are robust to a number of alternative specifications. The main results for fine particulate matter, low birth weight and infant deaths are shown in row 1 for reference. In row 2, we cluster the standard errors at the state-level to address the possibility of spatial dependence in the data and find no consequential change in precision. The main results limit the sample to counties whose centroids are within 200km of heavy ship traffic because counties far from the coast are less likely to provide suitable counterfactuals. We show that our results are robust to alternative choices for inclusion in the sample. Rows 3 and 4 of Table A5 show very similar estimates when we limit the sample to counties within 150km or 300km as well.

While our main specification includes region-by-year fixed effects, we show that the results are robust to more flexible state-by-year fixed effects in row 5. Row 6 includes more flexible weather controls. For each weather variable, we include 7 bins: below 5th percentile, 5 bins for even intervals from the 5th to 95th percentile, and above the 95th percentile. Next, rows 7-8 relax the balanced panel requirement for air quality monitors. Rather than restricting the sample to

[28] We obtain the point-locations of principal ports, as defined by the US Army Corps of Engineers, from the National Oceanic and Atmospheric Administration (NOAA, 2023). For purposes of comparison, we use the 27 major ports for ocean-going vessels as defined in Gillingham and Huang (2021).

balanced monitors from 2008 to 2016, row 7 only requires balance between 2009 and 2014. This increases our sample of counties from 232 to 251. Row 8 relaxes the requirement for a sample of balanced monitors and reports the unbalanced panel results. In row 9, we use an alternate measure of intensity of treatment that is based on the CMAQ prediction of total emissions from maritime shipping.

Next, we address concerns that other pollution abatement policies may occur during our sample period. First, we exclude counties with a port in row 10 to show our results are not driven by any port-specific policy changes that may have been adopted during our sample period. Our results are not driven by port counties alone. Second, row 11 shows our results are robust to controlling for Clean Air Act non-attainment status for each county over time. Third, we consider that the ECA policy we study also tightened standards for engine emissions of nitrogen oxides for a small subset of ship traffic. As an additional check to isolate the effect of the ECA on fuel standards, we show that the reduced form effects of the ECA on PM2.5, low infant birth weight, and infant death are robust to including NO2 as a control variable. For each of these robustness exercises in rows 2 through 12, the estimates remain significant and are similar in magnitude across each outcome.

Finally, row 13 tests whether the tightening of the fuel content standard in 2015 had any additional impact on improving air quality. We find no statistically significant impact on air quality or health outcomes from this tightening. This is not surprising, as the 2015 fuel standard tightening was a relatively small change and many ships were already using compliant fuel.

Table A5: **Robustness of Main Results**

	PM2.5		Low BW		Death <1		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	β	p-value	β	p-value	β	p-value	N clusters
(1) Baseline	-0.55	0.00	-1.33	0.00	-0.24	0.01	232.0
	(0.10)		(0.35)		(0.09)		
(2) State-level clustering	-0.55	0.00	-1.33	0.01	-0.24	0.00	25.0
	(0.11)		(0.43)		(0.05)		
(3) 150 km	-0.57	0.00	-1.32	0.00	-0.21	0.02	202.0
	(0.11)		(0.38)		(0.09)		
(4) 300 km	-0.56	0.00	-1.28	0.00	-0.25	0.01	280.0
	(0.10)		(0.32)		(0.09)		
(5) State-year FE	-0.45	0.00	-1.07	0.00	-0.21	0.03	232.0
	(0.13)		(0.32)		(0.10)		
(6) Bins of weather	-0.55	0.00	-1.37	0.00	-0.23	0.01	232.0
	(0.10)		(0.32)		(0.09)		
(7) 2009-2014 balance	-0.61	0.00	-1.38	0.00	-0.17	0.02	251.0
	(0.12)		(0.43)		(0.07)		
(8) Unbalanced panel	-0.39	0.00	-1.25	0.00	-0.24	0.00	286.0
	(0.10)		(0.31)		(0.08)		
(9) Ships' contribution	-0.38	0.00	-1.10	0.00	-0.20	0.00	232.0
	(0.11)		(0.28)		(0.07)		
(10) No ports	-0.58	0.00	-1.39	0.06	-0.51	0.00	192.0
	(0.18)		(0.74)		(0.17)		
(11) CAA controls	-0.42	0.00	-1.27	0.00	-0.24	0.02	232.0
	(0.11)		(0.34)		(0.10)		
(12) NO2 controls	-0.25	0.04	-0.75	0.01	-0.27	0.02	79.0
	(0.12)		(0.28)		(0.11)		
(13) 2015 0.1ppm	-0.02	0.82	0.20	0.65	-0.04	0.68	232.0
	(0.10)		(0.45)		(0.10)		

Note: Row 1 replicates the baseline results for PM2.5, low birth weight, and infant deaths from Panel A of Table 3. Columns 1, 3, and 5 report coefficients from estimating equation 1. Columns 2, 4, and 6 report p-values. Column 7 reports total number of clusters in each specification. Robust standard errors clustered at the county level are reported in parentheses, unless otherwise noted. Rows 2-11 present robustness checks. Row 2 clusters standard errors at the state level. Rows 3 and 4 limit the sample of counties to those with population-weighted centroids within 150km and 300km of heavy ship traffic, respectively. Row 5 replaces region-by-year fixed effects with state-by-year fixed effects. Row 6 includes more flexible binned weather controls. For each weather variable, we include 7 bins: below 5th percentile, 5 bins for even intervals from the 5th to 95th percentile, and above the 95th percentile. Rows 7-8 relax the balanced panel requirement for air quality monitors by restricting to a sample of balanced monitors from 2009 to 2014 (row 7) and to a sample of all counties that ever have PM2.5 data during the period of study (row 8). Row 9 examines the robustness of the treatment definition by employing the CMAQ prediction of total emissions from maritime shipping. Row 10 excludes counties with a port. Row 11 includes controls for Clean Air Act attainment status. Row 12 examines the effect of tightening the fuel content standard nationally in 2015.

B.7 Sensitivity to Fixed Effects and Continuous Treatment

We further scrutinize the baseline specification in Table A6. It shows our estimates across a variety of specifications for our three main outcomes in panels a-c: PM2.5, low birth weight, and infant mortality.

Estimates in column 1 include only county and year-by-month fixed effects. However, it is important to note that these estimates do not allow for differential regional trends and only control for nationally uniform time trends. If counties in California, for example, have substantially different trends or yearly shocks than counties in Texas in terms of pollution and/or health outcomes, these counties will not make a good counterfactual comparison. This seems likely. By controlling for only nationally uniform time trends in column 1, there are many potentially omitted variables at the region-by-year level that can bias estimates. This bias can be observed when we add our baseline vector of controls for weather, demographics, and unemployment rates in column 2. Comparing columns 1 and 2, we observe large changes in sign and magnitude for the estimated effects, suggesting large potential for bias from unobserved factors in this specification (Oster, 2019).

Next, we add region-by-year fixed effects in column 3. Including region-by-year fixed effects allows for differential trends or shocks by region and makes counterfactual comparisons only within the same region. Estimates in column 3 are statistically significant for all three outcomes of interest. This is not surprising, as trends in health and pollution will be more similar within the same region, and therefore these within-region comparisons provide better counterfactual comparisons. In Table A5 row 5, we also show our results are robust to including more granular state-by-year fixed effects as well.

Columns 4-7 show a variety of other specifications, for completeness. Across all these specifications, our estimates are similar in magnitude and statistically significant. Column 4 drops year-by-month fixed effects. Even without these controls, estimates in columns 3 and 4 are almost identical. In column 5, we include county-by-season fixed effects, rather than county fixed effects (including county fixed effects would be co-linear with county-by-season fixed effects). This allows for differential seasonal patterns of pollution and health for each county. As different counties experience different seasonal weather and pollution patterns, for example, we view this specification as capturing additional sources of bias at the county-season level. Column 6 adds our baseline vector of controls for weather, demographics, and unemployment, and is our preferred specification used in the paper. Comparing coefficients between columns 5 and 6 shows much smaller coefficient changes than the comparison of columns 1 and 2, suggesting a much more limited potential for unobserved factors to bias this preferred specification, which is reassuring (Oster, 2019). An even more fully saturated model is provided in column 7, which adds year-by-month fixed effects and the coefficients remain very similar. We find the similarity in our estimates across these specifications reassuring.

We also to provide a binary difference-in-difference model in column 8. In this specification, we interact our indicator for post-ECA with an indicator for the treatment group, counties above the weighted mean of predicted PM2.5 decline from the CMAQ model, 0.76. (Note that other cutoffs yield similar results.) This specification has the disadvantage that it does not leverage the continuous nature of our treatment, but the coefficients remain statistically significant across each of our outcomes. As expected, counties with greater exposure to the policy, with CMAQ estimates above 0.76, have a larger relative decline in PM2.5, low birth weight, and infant death. Fine particulate matter declines by 0.526 units in high exposure areas relative to low exposure areas after the policy, which is 5.04 percent relative to the pre-policy mean in high exposure areas, 10.45.

Table A6: **Role of Fixed Effects on Reduced Form Estimates**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. PM2.5								
Post-ECA*CMAQ	0.013	-0.261	-0.405	-0.395	-0.441	-0.554	-0.570	
	(0.117)	(0.134)*	(0.081)***	(0.081)***	(0.093)***	(0.104)***	(0.108)***	
Post-ECA*1(CMAQ > 0.76)								-0.526
								(0.259)**
R^2	0.52	0.68	0.53	0.40	0.52	0.60	0.69	0.59
N	24,901	24,901	24,901	24,901	24,901	24,901	24,901	24,901
County	X	X	X	X				
Year-by-month	X	X	X				X	
County-by-season					X	X	X	X
Region-by-year			X	X	X	X	X	X
X_{imy}		X				X	X	X
Panel B.Low birth weight								
Post-ECA*CMAQ	-0.465	-0.803	-1.344	-1.345	-1.349	-1.326	-1.308	
	(0.239)*	(0.260)***	(0.337)***	(0.336)***	(0.351)***	(0.348)***	(0.348)***	
Post-ECA*1(CMAQ > 0.76)								-1.760
								(0.481)***
R^2	0.55	0.56	0.55	0.54	0.56	0.57	0.57	0.57
N	25,052	25,052	25,052	25,052	25,052	25,052	25,052	25,052
County	X	X	X	X				
Year-by-month	X	X	X		X		X	
County-by-season					X	X	X	X
Region-by-year			X	X	X	X	X	X
X_{imy}		X				X	X	X
Panel C. Infant death								
Post-ECA*CMAQ	-0.073	-0.085	-0.254	-0.253	-0.254	-0.242	-0.246	
	(0.065)	(0.076)	(0.087)***	(0.087)***	(0.088)***	(0.089)***	(0.089)***	
Post-ECA*1(CMAQ > 0.76)								-0.442
								(0.165)***
R^2	0.62	0.62	0.62	0.62	0.63	0.63	0.63	0.63
N	25,052	25,052	25,052	25,052	25,052	25,052	25,052	25,048
County	X	X	X	X				
Year-by-month	X	X	X				X	
County-by-season					X	X	X	X
Region-by-year			X	X	X	X	X	X
X_{imy}		X				X	X	X

Note: The unit of observation is county-year-month. The sample includes counties with population-weighted centroids within 200km of heavy ship traffic and with a PM2.5 monitor from 2008-2016. The observations are weighted by the number of conceptions. The reduced-form estimates obtained from equation 1 are reported with varying fixed effects indicated in the column notes. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

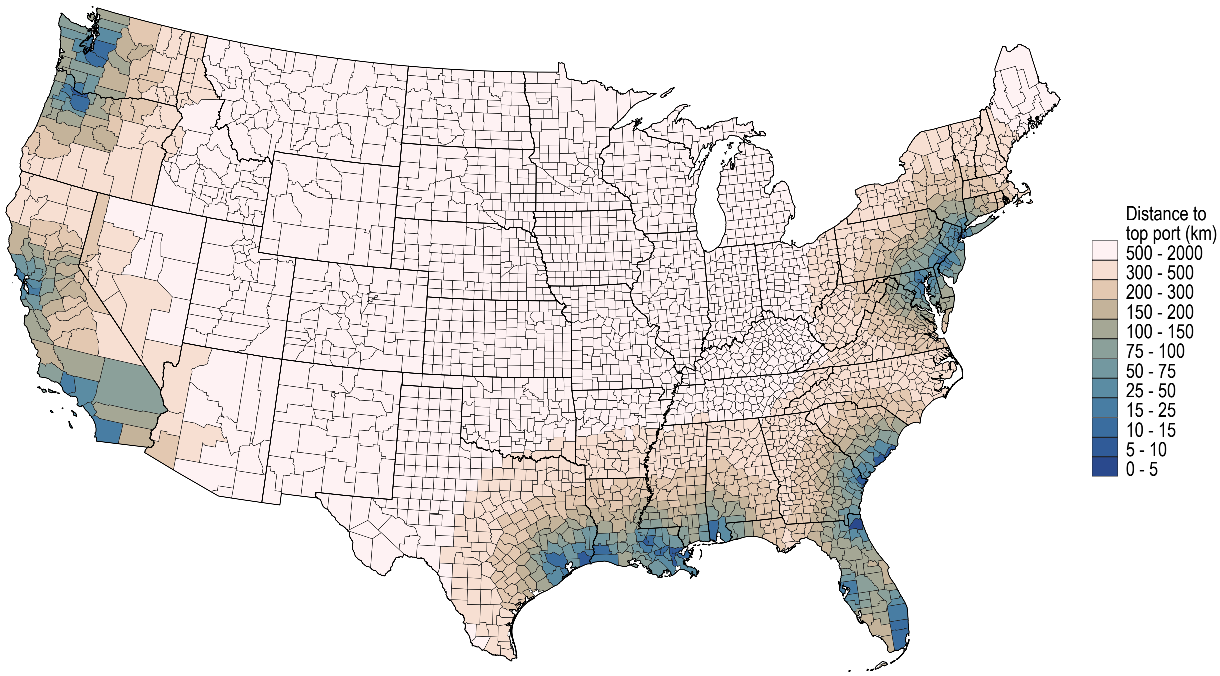
B.8 Comparing CMAQ with Distance Approach

Employing the CMAQ output as a measure of intensity of exposure to the ECA policy improves on approaches that rely on imprecise proxies for source-specific exposure. To illustrate the distinction, we perform our analysis using distance to a port as the proxy for ship pollution exposure in lieu of CMAQ output. We define distance to a port as the kilometers from the county population-weighted centroid to the nearest major US port (shown in Figure A7).[29] We highlight two main concerns with distance metrics. First, distance is a poor proxy for exposure to improvements from the policy because atmospheric interactions play a major role in the dispersion of pollution. This concern would lead to bias from measurement error. Second, there does not exist an *a priori* functional form for the relationship between the distance from a pollution source and pollutant exposure from the source. This concern would lead to bias from misspecification.

Table A7 reports the results of estimating equation (1) where intensity of exposure to the policy is measured by either CMAQ or distance. We report the results for infant deaths, low birth weight, and fine particulate matter in panels A-C, respectively. We standardize the coefficients and standard errors into units of standard deviations so that the results are comparable across candidate treatment variables. First, we compare the Bayesian information criteria (BIC), which is a criterion for model selection based, in part, on the likelihood function. We observe that the estimates based on CMAQ consistently yield a lower BIC across all outcomes, suggesting that CMAQ is preferred. Across all outcomes, the CMAQ model appears to reduce measurement error, as expected. The T-statistic is larger and standard errors are smaller for CMAQ relative to distance in all panels. The estimated effect of a one standard deviation increase in distance relative to CMAQ exposure led to a slightly larger reduction in infant deaths and low birth weight, but a slightly smaller reduction in fine particulate matter. However, we do not emphasize these differences because the confidence intervals of these estimates overlap. Nevertheless, these models show meaningful improvements in precision when CMAQ is used to measure exposure to the ECA policy.

[29] We obtain the point-locations of principal ports, as defined by the US Army Corps of Engineers, from the National Oceanic and Atmospheric Administration. For purposes of comparison, we use the 27 major ports for ocean-going vessels as defined in Gillingham and Huang (2021).

Figure A7: Distance to Ports



Note: Figure shows the distance from the population-weighted centroid of each county to the nearest principal port.

Table A7: **Comparison of Treatment Variables on Main Outcomes**

	BIC	T-stat	Coefficient	Std error
Panel A: PM2.5				
CMAQ	107,538.141	-5.346	-0.058	0.011
-Distance port	107,672.391	-1.665	-0.045	0.027
Panel B: Low birth weight				
CMAQ	190,136.859	-3.807	-0.015	0.004
-Distance port	190,155.312	-2.177	-0.019	0.009
Panel C: Infant Deaths				
CMAQ	136,256.172	-2.723	-0.010	0.004
-Distance port	136,259.469	-1.706	-0.017	0.010

Note: Table reports results of estimating equation 1 where the intensity of exposure to the policy is measured by either the CMAQ prediction or distance to the nearest major port. County-year-months are weighted by the number of conceptions. We report the results for fine particulate matter, low birth weight, and infant deaths in panels A-C, respectively. Coefficients and standard errors are standardized into units of standard deviations so that the results are comparable across candidate treatment variables. Column 1 reports the Bayesian information criteria (BIC), where the lowest BIC is preferred. Columns 2-4 report the T-statistic, coefficient, and standard error, respectively.

C Incidence Results

C.1 Exposure Across Demographic Groups Results

Despite existing evidence of disproportionate pollutant exposure for disadvantaged groups from land-based pollution sources, prior work has not examined the exposure gap for a source that is mobile and at-sea. We highlight the differences in the demographics of the population exposed 13 to maritime fuel emissions versus comparable stationary on-land sources in Figure A8. Figure A8 shows the correlations between race/ethnicity and two measures of exposure to maritime pollution: distance to ports (stationary on-land) and the overall intensity of ship emissions, as measured by the CMAQ model (mobile at-sea).[30] We report results for non-Hispanic white, non-Hispanic black, non-Hispanic other race, and Hispanic. We use demographic information from 2010 census tract data (U.S. Census Bureau, 2010). We restrict our sample for analysis to tracts within 200km of heavy ship traffic. Each of the 100 circles represents the population-weighted average for equal-sized bins of census tracts.

First, panels (a)-(d) show the relationship between race/ethnicity and distance to ports. We calculate the distance from the population-weighted centroid of each tract to the nearest large port. Counties further to the right are closer in distance to a port. Consistent with the environmental justice literature, the population near ports is less likely to be white (panel (a)), and more likely to be black, other race, or Hispanic (panels (b)-(d)).

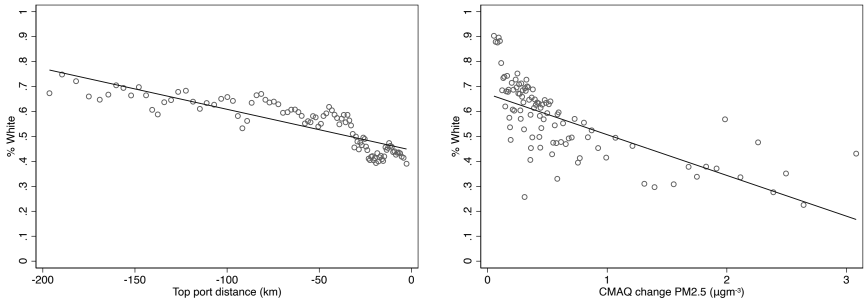
While ports are an important source of air pollution, exposure to maritime pollution from shipping routes is not captured by the distance-to-port measure. To account for the total contribution of ship emissions to a local area's pollution levels, panels (e)-(h) show the correlation between race and intensity of ship emissions, as measured by the predicted change from requiring low-sulfur maritime fuel, based on the CMAQ model. The x-axis reports the predicted change in fine particulate matter. Census tracts further to the right are predicted to have larger improvements in air quality from the maritime fuel regulation. Interestingly, the correlation between the proportion of non-Hispanic black individuals and maritime emissions intensity shown in panel (f) is negative. This pattern is in contrast to most other pollution contexts, including distance to ports. All other race/ethnicity groups show correlations in the same direction as those observed for distance to ports. However, the slopes of each differ somewhat, especially for Hispanics.

Figure 8 shows an alternative way to visualize these patterns. Here we present the cumulative distribution function of the proportion of individuals in each race/ethnicity group over distance to port (panel (a)) and intensity of ship emissions (panel (b)). A few interesting patterns stand out and are consistent with Figure A8. First, panel (a) shows that non-Hispanic blacks are more likely to live very near ports. In general, non-white individuals are more likely to live near ports and non-Hispanic whites are least likely to live near ports, consistent with the large environmental justice literature looking at stationary land-based pollution sources. However, the pattern is different in panel (b), which shows the cumulative distribution of individuals by intensity of exposure to overall ship emissions, as measured by CMAQ. Unlike panel (a), black and white individuals have almost identical distributions, suggesting they experience a much more similar distribution of exposure to overall ship emissions. Moreover, both groups are less likely to live with high exposure to ship emissions, relative to Hispanics and non-Hispanic other race groups.

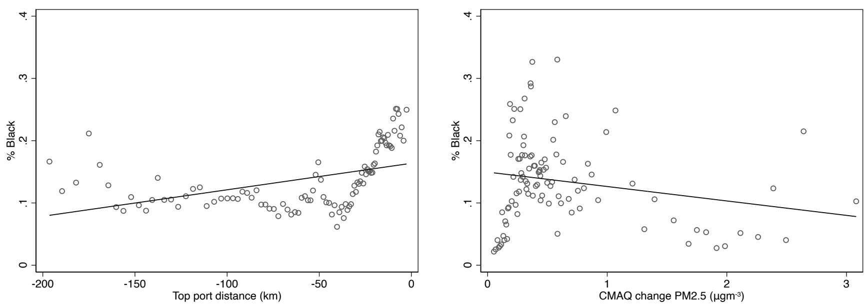
[30] Figure A7 shows the distance to ports, and Figure 3 shows the overall intensity of ship emissions based on the CMAQ model.

Given that the exposed population is different from most land-based pollution sources, the health effects of this policy are likely to be different than other reductions in air pollution. This is likely to be the case if, for example, pollution has a heterogeneous health impact across demographic groups, perhaps due to differences in underlying health conditions or access to care. In addition, the dose of exposure to maritime pollution may differ across demographic groups due to differences in time spent outdoors or differential avoidance behaviors. We compare the magnitude of our health results to the health effects of pollution found in other contexts to better understand the extent to which these differences in the demographics of individuals exposed to maritime emissions yield different overall effects on health.

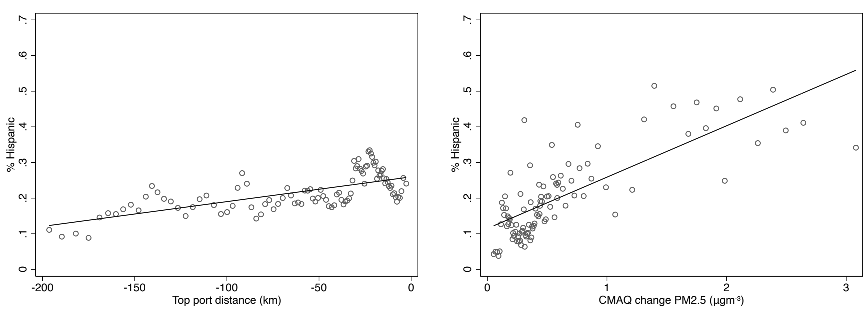
Figure A8: **Demographics of the Population Exposed to Maritime Pollution**



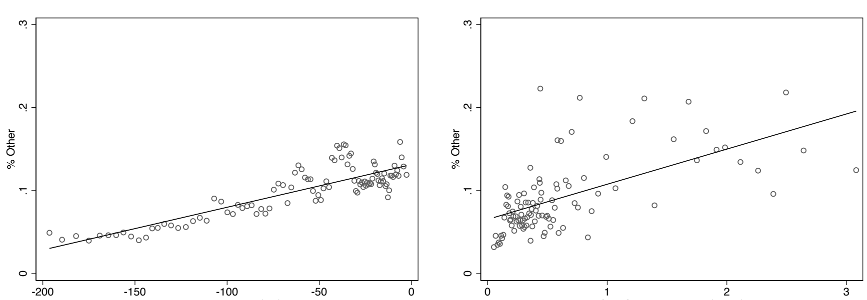
(a) Distance to Port & NH White (e) CMAQ Exposure & NH White



(b) Distance to Port & NH Black (f) CMAQ Exposure & NH Black



(c) Distance to Port & Hispanic (g) CMAQ Exposure & Hispanic



(d) Distance to Port & NH Other (h) CMAQ Exposure & NH Other

Note: Demographic information on the proportion of non-Hispanic whites, non-Hispanic blacks, non-Hispanic other race, and Hispanics from 2010 census tract data. We restrict to our analysis sample, which includes tracts within 200km of heavy ship traffic. Each of the 100 circles represents the population-weighted average for equal-sized bins of census tracts. Panels (a)-(d) show the correlation between race/ethnicity groups and distance to ports. We calculate distance from the population-weighted centroid of each tract to the nearest major port. Counties further to the right are closer in distance to a port. Panels (e)-(h) show the correlation between race/ethnicity and the intensity of ship emissions, as measured by the predicted change from requiring low sulfur maritime fuel from the CMAQ model at the centroid of each tract. The x-axis is the predicted change in fine particulate matter. Counties further to the right have higher ship emission exposure.

C.2 Incidence across Demographic Groups Results

Table A8: **Heterogeneity of Effects of PM2.5 from ECA on Low Birth Weight**

	(1) All	(2) NH White	(3) NH Black	(4) NH Other	(5) Hispanic	(6) High Educ	(7) Married	(8) Age 19-24	(9) Age 25-34	(10) Age 35
PM2.5	0.00266	0.00301	0.00212	0.00756	0.00107	0.00225	0.00181	0.00106	0.00266	0.00417
	(0.00093)***	(0.00153)*	(0.00256)	(0.00209)***	(0.00060)*	(0.00106)**	(0.00059)***	(0.00087)	(0.00106)**	(0.00120)***
R^2	0.01	0.01	0.01	-0.00	0.00	0.01	0.01	0.01	0.01	0.01
N	12,426,807	5,062,128	1,860,002	1,337,613	4,167,051	6,436,488	7,238,190	2,911,813	6,967,317	2,317,026
F	21.44	11.38	11.62	26.84	24.33	19.25	19.98	19.02	20.06	27.37
N-counties	232	232	232	231	232	232	232	232	232	232
Mean	0.06	0.05	0.11	0.06	0.06	0.05	0.05	0.07	0.06	0.07
%Change	4.39	6.42	1.99	12.03	1.92	4.17	3.63	1.57	4.81	6.42

Note: The unit of observation is the individual-year-month. The sample includes individuals in counties with population-weighted centroids within 200km of heavy ship traffic and with a PM2.5 monitor from 2008-2016. Observations are unweighted. Results show two-stage least squares estimates based on equation 2. Column 1 includes the entire sample and reports results analogous to Table 3 panel B, column 2, but at the individual level. Columns 2-10 restrict the sample to individuals of different demographic groups, including non-Hispanic white, non-Hispanic black, non-Hispanic other, Hispanic, highly educated, married, age 19-24, age 25-34, and age 35+. Within county-season R^2 is reported. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

D Behavioral Response Additional Results

D.1 Ship Behavior

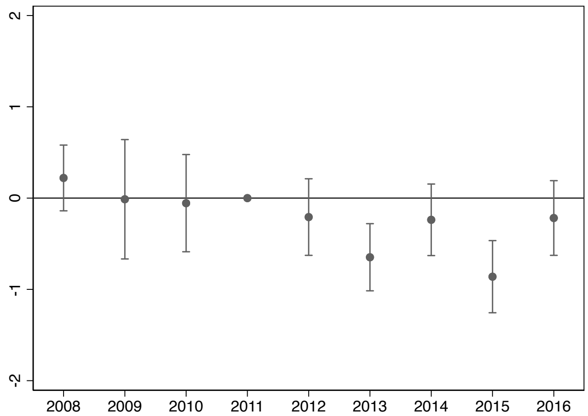
In this section, we explore the year-to-year variation in the effect of the policy. We found the variation was driven by the counties in southern California with partial-ECA coverage.[31] To establish this, Figure A9 presents the effect of the ECA PM2.5 separately for southern CA (panel a) and elsewhere (panel b).

To further understand this pattern, we repeated our analysis with the addition of southern California-by-year fixed effects to allow for differential yearly trends in southern California. Panel (c) of Figure A9 repeats panel (a) with these additional controls and panel (d) repeats our baseline result for the full sample with these additional controls. Once these controls are included, for both groups the "bouncy" pattern observed in the post period is eliminated while the decline in pollution remains statistically significant and similar in magnitude to our main estimates. This indicates there were local shocks, perhaps to weather or pollution, in southern California that were not perfectly captured by our baseline control variables; however, the addition of more granular controls confirms a robust effect of the ECA policy and mitigates the year-to-year variation.

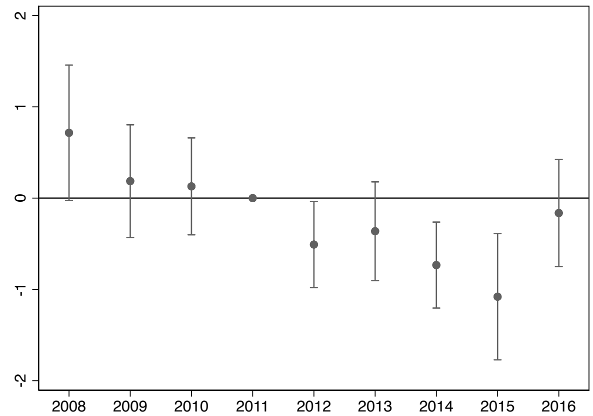
[31] We define southern California as the partial ECA counties in California. This includes nine counties: Imperial, Kern, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura.

Figure A9: **Influence of southern California on year-to-year variation**

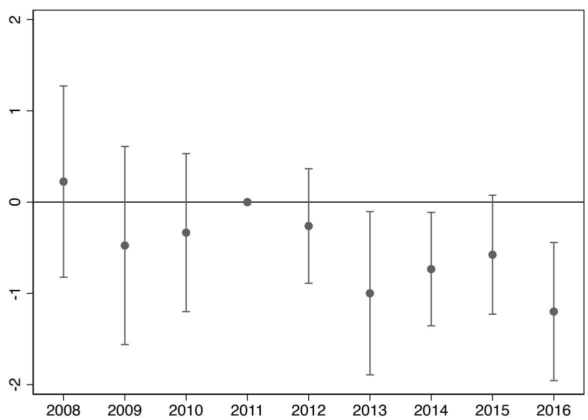
(a) Southern CA with baseline controls



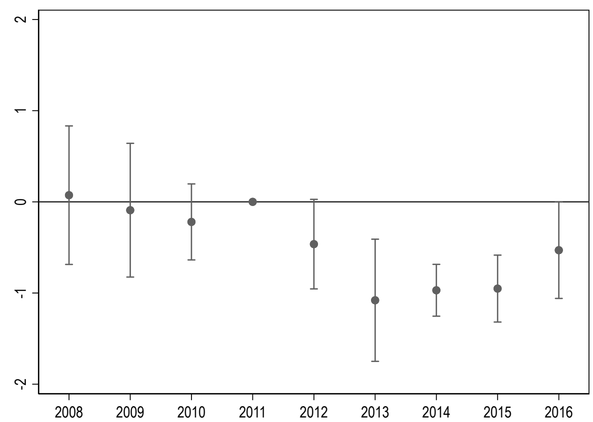
(b) Outside southern CA with baseline controls



(c) Southern CA with additional controls

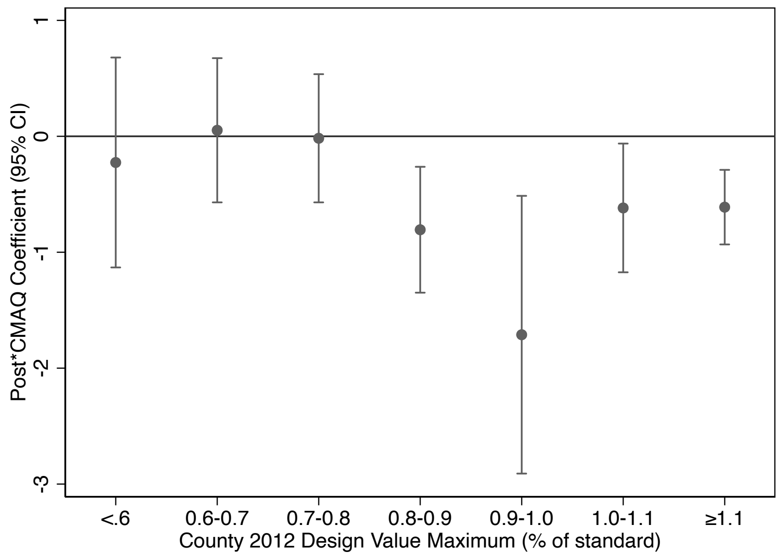


(d) Full sample with additional controls



D.2 Other Emissions

Figure A10: **Emissions Behavioral Response: Clean Air Act Regulatory Rebound**



Note: The outcome is fine particulate matter. The unit of observation is a county-year-month. The observations are unweighted. The sample includes counties with population-weighted centroids within 200km of heavy ship traffic and with a PM2.5 monitor. The depicted coefficients are the estimated effect of a one-unit increase in a county's CMAQ predicted reduction from the ECA in post-ECA (July 2012) time periods relative to pre-ECA time periods. Robust standard errors are clustered at the county level. The confidence intervals are $\pm$ 1.96 standard errors.

Table A9: **Effect of ECA on Ship and Other Emissions Behavior**

	(1) PM2.5	(2) PM2.5	(3) PM2.5
Post*CMAQ	-0.610	-0.860	-1.713
	(0.129)***	(0.195)***	(0.602)***
Post*CMAQ*1(ECA<200nm)		0.372	
		(0.176)**	
Post*CMAQ*1(DV<0.8)			1.685
			(0.560)***
Post*CMAQ*1($0.8 \leq$ DV < 0.9)			0.906
			(0.590)
Post*CMAQ*1($1.0 \leq$ DV)			1.102
			(0.578)*
R^2	0.62	0.62	0.58
N	24,905	24,905	19,992
N-counties	232	232	186
Mean	8.30	8.30	8.72
% Change:			
All	-7.35		
ECA=200nm		-10.41	
ECA<200nm		-5.71	
DV < 0.8			-0.36
$0.8 \leq$ DV < 0.9			-9.60
$0.9 \leq$ DV < 1.0			-17.00
$1.0 \leq$ DV			-6.42

Note: The unit of observation is the county-year-month. The sample includes counties with population-weighted centroids within 200km of heavy ship traffic and with a PM2.5 monitor from 2008-2016. County-year-months are not weighted. Column 1 shows the results of estimating equation 1. Column 1 repeats the estimate of Table 2 column 1 with unweighted data. Column 2 repeats column 1 with an additional interaction for whether the ECA boundary is less than the full 200 nm from the county population-weighted centroid, 1(ECA<200 nm), as per equation 3. Column 3 repeats column 1 with additional interactions for counties' pre-policy distance to the regulatory threshold, DV, defined as the county 2012 PM2.5 maximum design value as a percent of the standard, as per equation 4. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

In addition to results presented in Figure A10 and Table A9, we present additional empirical evidence consistent with this hypothesis in Table A10. To present this evidence, we simplify the model to focus on counties with DV above and below 0.8. Our previous results estimated a more flexible model, but the main finding was that the effect of the ECA policy on air quality was muted for counties with DV <0.8, consistent with regulatory rebound in these counties. This streamlined model preserves power for the additional analyses we present below.

We begin with additional evidence consistent with our conceptual intuition that counties with DV<0.8 are reasonably constrained by the NAAQS. As anecdotal evidence that counties with DV<0.8 are bound by the NAAQS, we note that of the 53 counties with DV <0.8, half are either in non-attainment for another pollutant or were previously in non-attainment for any pollutant. This provides an additional way to measure which counties are relatively more constrained under the NAAQS. We start by exploring heterogeneity by previous non-attainment or non-attainment for another pollutant. In columns 3 and 4, we show below and above 0.8 DV counties, respectively. For low DV counties in column 3, we see that the counties with past non-compliance or non-compliance for other pollutants are largely contributing to the rebound among counties with DV <0.8 because their post-ECA declines in PM2.5 per unit CMAQ are not different than zero. By contrast, for DV$\geq$0.8 counties in column 4, all counties demonstrate comparable decline in ambient PM2.5 regardless of regulatory status.

Finally, we look for evidence of emissions changes that would yield the patterns we observe in ambient PM2.5. Table A10 columns 5-6 explore heterogeneity in the effect of the policy on emissions data from the EPA's National Emissions Inventory (NEI). Column 5 reports the effect of the ECA on the log of PM2.5 emissions while column 6 reports the effect on the log of PM1 (ultrafine particles) emissions. For both pollutants, we found that the magnitude of the effect on emissions was larger for counties with DV$<$0.8 than in counties with DV$\geq$0.8, consistent with our hypothesis and the effects on ambient PM2.5, but these differences are not statistically significant. Overall, the emissions results in columns 5-6 suggest increasing emissions may have been a response to the ECA policy, but we cannot rule out other types of strategic response that we do not observe.

Table A10: **Clean Air Act heterogeneity in ambient PM2.5 and total emissions**

	(1) PM2.5	(2) PM2.5	(3) PM2.5 DV<0.8	(4) PM2.5 DV>=0.8	(5) log(PM2.5) (NEI)	(6) log(PM1) (NEI)
Post*CMAQ	-0.610	-0.470	-0.527	-0.899	-0.039	0.004
	(0.155)***	(0.233)**	(0.469)	(0.392)**	(0.066)	(0.108)
Post*CMAQ*1(DV<0.8)	0.715				0.108	0.196
	(0.200)***				(0.140)	(0.235)
Post*CMAQ*1(past or other NA)		0.088	0.581	0.107		
		(0.240)	(0.363)	(0.405)		
Post*CMAQ*1(2012 PM2.5 NA)		-0.168		0.365		
		(0.219)		(0.372)		
R^2	0.58	0.62	0.57	0.57	0.92	0.89
N	19,992	24,905	5,685	14,307	6,696	6,420
N-counties	186	232	53	133	186	184
Mean	8.72	8.30	7.55	9.19	5.48	4.96
% Change:						
DV<0.8	1.40				6.91	20.00
DV>=0.8	-6.63				-3.89	0.39
Always Attain		-6.07	-6.90	-9.75		
Past or other NA		-4.78	0.72	-9.07		
2012 PM2.5 NA		-6.71		-5.61		

Note: The unit of observation is county-year-month. The sample includes counties with population-weighted centroids within 200km of heavy ship traffic and with a PM2.5 monitor from 2008-2016. County-year-months are not weighted. In columns (1) and (5)-(6), estimates are based on equation 1 with an additional interaction of the main variable with an indicator for below 80 percent of the regulatory design value threshold. In columns (2)-(4), estimates are based on equation 1 with an additional interaction of the main variable with a categorical variable for a county's compliance history with the Clean Air Act: {always attainment, past non-attainment or 2012 non-attainment with other pollutant, 2012 non-attainment with PM2.5}. In columns (5) and (6), emissions are measured as the log of the county annual sum of fine particulate matter (column 5) and ultrafine particulate matter (column 6) in the National Emissions Inventory for 2008, 2011, 2014, and 2017. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

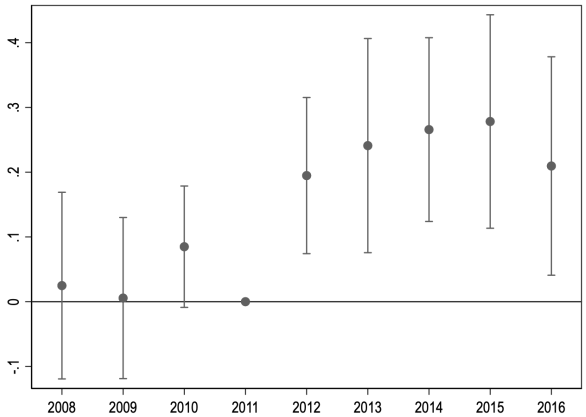
D.3 Individual Behavior

Table A11: **Effect of ECA on Individual Behavior**

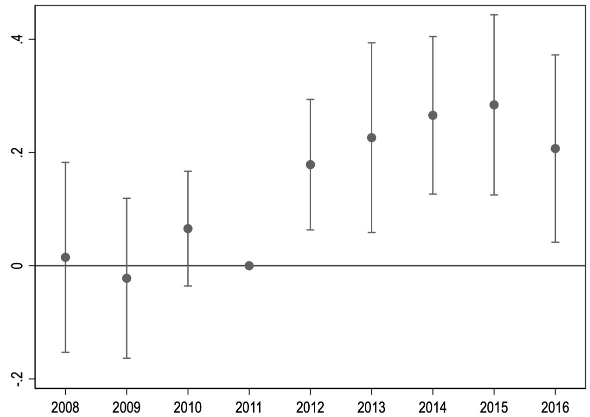
	Campsite Reservations (IHS)						Time Outdoors
	Visits	Visits	People	People	Days	Days	(IHS)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
post-ECA × CMAQ	0.164***	0.144***	0.149***	0.144***	0.166***	0.147***	0.0797*
	(0.0603)	(0.0325)	(0.0480)	(0.0357)	(0.0597)	(0.0307)	(0.0473)
Region-year FE	X	X	X	X	X	X	X
County-season FE	X		X		X		X
Facility-month FE		X		X		X	
Year-month FE		X		X		X	X
R-squared	0.879	0.944	0.871	0.927	0.906	0.950	0.064
Observations	10,909	38,385	10,909	38,385	10,909	38,385	29,516
N-counties	158	150	158	150	158	150	183
Mean	437	124	2,210	626	1,178	334	14.68

Note: The unit of observation is county-year-month in columns 1, 3, and 5, and is facility-year-month for columns 2, 4, and 6. For columns 1-6, observations are unweighted, and the sample is counties with population-weighted centroids within 200km of heavy ship traffic. In columns 1-6, we estimate equation 5 where the outcomes are the inverse hyperbolic sine of the number of visits (columns 1-2), the number of people (columns 3-4), and the number of days (columns 5-6). In column 7, the unit of observation is the individual-year-month, observations are weighted with survey weights, and the sample includes observations in counties with population-weighted centroids within 200km of heavy ship traffic. In column 7, we estimate equation 6 where the outcome is the inverse hyperbolic sine of the number of minutes the respondent reported spending outdoors for the previous day. All regressions include region-by-year fixed effects; columns 1, 3, 5, and 7 include county-by-season fixed effects; columns 2, 4, and 6 include facility-by-month fixed effects; and columns 2, 4, 6, and 7 include year-by-month fixed effects. Column 7 also controls for gender, race, ethnicity, education, age, presence of children in the household, and indicators for the day of the week of the survey and whether it was a holiday. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Figure A11: **Individual Behavioral Response: Campsite Reservations**



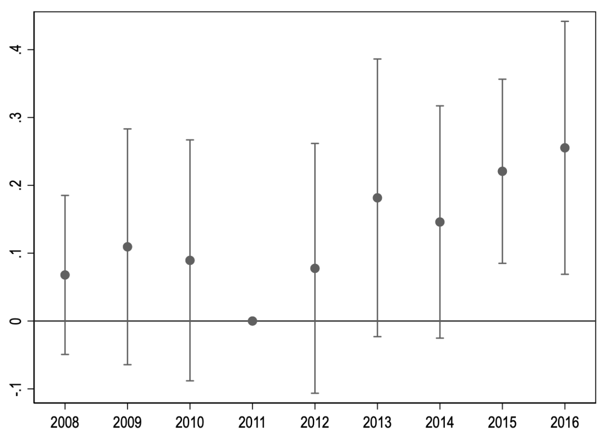
(a) Visits



(b) Days

Note: The unit of observation is the county-year-month. The observations are unweighted. The sample includes counties with population-weighted centroids within 200km of heavy ship traffic. The depicted coefficients are the estimated effect of a one-unit increase in a county's CMAQ predicted reduction from the ECA in each year relative to 2011, the year prior to policy adoption. Panels a and b show the estimates for outcomes variables (a) inverse hyperbolic sine of total visits and (b) total days each month, respectively. Robust standard errors are clustered at the county level. The confidence intervals are $\pm$ 1.96 standard errors.

Figure A12: **Individual Behavioral Response: Time Spent Outdoors**



Note: The unit of observation is the individual-county-year-month. The observations are weighted by sample weights. The sample includes individuals in counties with population-weighted centroids within 200km of heavy ship traffic and with a PM2.5 monitor. The depicted coefficients are the estimated effect of a one-unit increase in a county's CMAQ predicted reduction from the ECA in each year relative to 2011, the year prior to policy adoption. The outcome is the inverse hyperbolic sine transformation of minutes spent outdoors. Robust standard errors are clustered at the county level. The confidence intervals are $\pm$ 1.96 standard errors.

Table A12: **Effect of ECA on Campsite Reservations (Log)**

	Visits (1)	Visits (2)	People (3)	People (4)	Days (5)	Days (6)
post-ECA × CMAQ	0.168***	0.146***	0.149***	0.144***	0.173***	0.150***
	(0.0609)	(0.0335)	(0.0487)	(0.0362)	(0.0606)	(0.0310)
Region-year FE	X	X	X	X	X	X
County-season FE	X		X		X	
Facility-month FE		X		X		X
Year-month FE		X		X		X
R-squared	0.859	0.934	0.847	0.899	0.855	0.933
Observations	10,508	37,374	10,508	37,373	10,094	35,811
N-counties	148	143	148	143	140	135
Mean	454	127	2,294	643	1,273	358

Note: The unit of observation is county-year-month in columns 1, 3, and 5, and is facility-year-month for columns 2, 4, and 6. Observations are unweighted, and the sample is counties with population-weighted centroids within 200km of heavy ship traffic. We estimate equation 5 where the outcomes are the natural log of the number of visits (columns 1-2), the number of people (columns 3-4), and the number of days (columns 5-6). All regressions include region-by-year fixed effects; columns 1, 3, and 5 include county-by-season fixed effects; columns 2, 4, and 6 include facility-by-month fixed effects; and columns 2, 4, and 6 include year-by-month fixed effects. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A13: **Effect of ECA on Time Spent Outdoors: Extensive and Intensive Margins**

	IHS (1)	Any (2)	Log (3)
post-ECA × CMAQ	0.0797*	0.0156	0.0553
	(0.0473)	(0.00986)	(0.0663)
County-season FE	X	X	X
Region-year FE	X	X	X
Year-month FE	X	X	X
R-squared	0.064	0.065	0.241
Observations	29,516	29,516	5,033
N-counties	183	183	162
Mean	14.68	0.174	84.81

Note: The unit of observation is the individual-year-month, observations are weighted with survey weights, and the sample includes observations in counties with population-weighted centroids within 200km of heavy ship traffic. We estimate equation 6 where the outcome is a measure of minutes the respondent reported spending outdoors for the previous day: column 1 uses the inverse hyperbolic sine of minutes outdoors, column 2 uses an indicator for any minutes outdoors, and column 3 uses the log of minutes outdoors, excluding zeros. All regressions include region-by-year fixed effects, county-by-season fixed effects, and year-by-month fixed effects. Regressions include controls for gender, race, ethnicity, education, age, presence of children in the household, and indicators for the day of the week of the survey and whether it was a holiday. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A14: **Effect of ECA on Bins of Time Spent Outdoors**

	0 hrs (1)	0-1 hrs (2)	1-2 hrs (3)	2-3 hrs (4)	3-5 hrs (5)	>5 hrs (6)
post-ECA × CMAQ	-0.0156	0.0141*	0.00307	-0.00319	-0.00270*	0.00434**
	(0.00986)	(0.00832)	(0.00334)	(0.00214)	(0.00158)	(0.00212)
R-squared	0.065	0.057	0.044	0.041	0.052	0.059
Observations	29,516	29,516	29,516	29,516	29,516	29,516
N-counties	183	183	183	183	183	183

Note: The unit of observation is the individual-year-month, observations are weighted with survey weights, and the sample includes observations in counties with population-weighted centroids within 200km of heavy ship traffic. We estimate equation 6 where the outcome is bins of the number of minutes the respondent reported spending outdoors for the previous day. All regressions include region-by-year fixed effects, county-by-season fixed effects, and year-by-month fixed effects. Regressions include controls for gender, race, ethnicity, education, age, presence of children in the household, and indicators for the day of the week of the survey and whether it was a holiday. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A15: **Time Outdoors: Placebo Tests**

	(1) Sleep	(2) Housework	(3) Groceries
post-ECA × CMAQ	0.00112	0.0486	-0.0235
	(0.00669)	(0.0691)	(0.0301)
Region-year FE	X	X	X
County-season FE	X	X	X
Year-month FE	X	X	X
R-squared	0.083	0.153	0.063
Observations	29,516	29,516	29,516
N-counties	183	183	183

Note: The regression specifications are identical to those in Table A11, but for the following outcomes: time spent sleeping (activity code 010101), time spent doing housework (activity codes 020101-020199), and time grocery shopping (activity code 070101).