Uncharted Waters: Effects of Maritime Emission Regulation

Jamie Hansen-Lewis* Michelle Marcus†

April 2023

Abstract

Maritime shipping emits as much fine particulate matter as half of global road traffic. We are the first to measure the consequences of US maritime emissions standards on air quality, human health, racial exposure disparities, and behavior. The introduction of US maritime emissions control areas significantly decreased fine particulate matter, low birth weight, and infant mortality. Yet, only about half of the forecasted fine particulate matter abatement was achieved by the policy. We show evidence consistent with behavioral responses among ship operators, other polluters, and individuals that muted the policy's impact, but were not incorporated in ex-ante models.

Over 80 percent of the volume of international trade is conducted via ship (UNCTAD, 2022), yet historical international standards for ship exhaust are strikingly weak in comparison to standards for other forms of transport that occur close to populated areas. For example, in 2008, the maximum allowable sulfur content of marine fuel along US coastlines was 3,500 times higher than that allowed in vehicles. Pollution from ship exhaust is a main component of poor air quality, not

*Hansen-Lewis is Assistant Professor in the Department of Agricultural and Resource Economics at the University of California, Davis (email: jhansenlewis@ucdavis.edu).

†Marcus is Assistant Professor in the Department of Economics at Vanderbilt University and a Faculty Research Fellow at the National Bureau of Economic Research (email: michelle.marcus@vanderbilt.edu). For their helpful comments, we thank Anna Aizer, Timothy Beatty, Fiona Burlig, Christopher Carpenter, Chloe East, Elaine Hill, Katrina Jessoe, Sabrina Eisenbarth, Emily Lawler, Casey Wichman, Jesse Shapiro, Andrew Foster, participants at the National Bureau of Economic Research 2022 Summer Institute, Association of Environmental and Resource Economists annual summer conference, the Southern Economic Association annual meeting and the American Society of Health Economics annual meeting, and seminar attendees at UC Davis, Brown University, Vanderbilt University, Vanderbilt University Medical Center, University of Memphis, Georgia Tech, Iowa State University, Rensselaer Polytechnic Institute, and Louisiana State University. We gratefully acknowledge financial support for this project from the Transportation Economics in the 21st Century Initiative of the National Bureau of Economic Research and the US Department of Transportation as well as the USDA National Institute of Food and Agriculture, Hatch project 1024474, and the Giannini Foundation of Agricultural Economics. All remaining errors are our own.

only at ports, but also in coastal communities near ship routes.[1] Since roughly half of the US population lives within 200 km of heavy ship traffic and ship traffic continues to increase, maritime emissions represent a significant threat to human and ecosystem health (U.S. EPA, 2009c,b, 2016). Yet, we lack a comprehensive understanding of the exposed population demographics and health effects of maritime emission regulation. Because ships are mobile and emissions occur off-shore, the health benefits from regulation are likely to differ from regulating land-based pollution sources.

Efficient design of maritime emission regulations is difficult for several reasons. First, the benefits to human health are uncertain. The health effects are likely to differ from regulation of other pollution sources because of the especially high sulfur content of ship fuel, the distinct coastal population exposed to ship traffic, and the degree to which individuals can avoid ship exhaust relative to other sources. Second, uniform regulation of maritime emissions along the coast will have heterogeneous effects due to the non-uniform population distribution and location of ships. The mobile nature of ship traffic makes it difficult to predict how ship routes and emissions may respond to regulation. Without comprehensive evidence accounting for spatial heterogeneity in the effects of maritime emission regulations on coastal population health, uniform regulation of maritime fuel risks both abating too little of emissions near coastal populations and too much of emissions far from coastal populations.

We provide the first evaluation of a major US environmental policy in the maritime shipping industry. In 2012, the US government, in coordination with the International Maritime Organization (IMO), introduced its seminal regulation of maritime pollution, called emission control areas (ECAs) (U.S. EPA, 2010). ECA regulation required all commercial ships to operate with low-sulfur fuel within 200 nautical miles off the coast or to install abatement equipment, or face penalties. In 2020, following this initial ECA regulation, the IMO extended similar standards globally, which were estimated would cost the shipping industry \$10 to \$60 billion per year (Corbett et al., 2016). Despite the consequential scale and cost of these regulations, ex-post evidence on the effectiveness and health benefits of such regulation has not been previously established.

[1] Roughly 70 percent of maritime emissions occur within 400 km of coasts, and maritime emissions elevate ambient fine particulate matter as much as 2 micrograms per cubic meter (Corbett et al., 2007). Maritime emissions account for roughly 38 percent of sulfur dioxide emissions on the US East Coast and 20 percent on the US West Coast (Wang et al., 2007). In areas adjacent to busy ports, they may equal or exceed those of land-based sources (Capaldo et al., 1999). Eyring et al. (2005) show maritime emissions are comparable to other transport modes. Smith et al. (2015) provide an emissions inventory.

In this paper, we measure the success of US maritime emissions standards and show evidence consistent with behavioral responses that diminished the policy's effectiveness relative to ex-ante predictions. With administrative data on air quality, infant health, and infant mortality, we use a differences-in-differences design and leverage variation in (i) the timing of the regulation and (ii) the intensity of the regulation across locations. Intensity of exposure to the policy is based on predictions from the Community Multiscale Air Quality Modeling System (CMAQ) obtained from U.S. EPA (2009c). CMAQ both represents policymakers' expectations of the effectiveness of the policy and provides a scientific prior of the policy's intended intensity at each location that accounts for atmospheric dispersion, disposition, and chemical interactions of pollution once emitted. After estimating the policy's ex-post benefits, we then test whether the changes in air quality from the policy were equal to the predicted improvements from the EPA's pre-policy analysis. We explain gaps between the ex-ante and ex-post predictions with behavioral responses on the part of the industry regulated by the policy, other pollution sources, and individuals, all of which can lessen the effectiveness of policy.

We find that the introduction of maritime emissions control areas around the US coastline led to a 4 percent decrease in the population-weighted average fine particulate matter across counties within 200km of heavy ship traffic.[2] We also find less of the disproportionate effects on minorities that have been documented as a result of land-based emissions, such as emissions from ports. Consistent with the air quality improvements, we find that the policy results in a 1.7 percent average reduction in the incidence of low birth weight. We also find a 2.8 percent decline in infant mortality. We further show that using an atmospheric aerosol transport model instead of distance as a proxy for exposure provided a meaningful improvement in estimation. Using distance in lieu of the CMAQ output would have yielded substantially less precision in the estimated effects of the policy.

These improvements in air quality and health demonstrate a substantial policy achievement. We estimate the ECA led to 1,536 fewer low birth weight infants and 228 fewer infant deaths per year. Our back-of-the-envelope calculation finds these improvements led to a savings of about $2.3 billion per year. These benefits to improved infant health alone are over 70 percent of the estimated

[2]Once primary pollutants such as sulfur exhaust are emitted in the atmosphere, they form secondary pollutants, such as particulate matter, through chemical interaction. Accordingly, the regulation defined fuel content limits for sulfur exhaust as a means to abate fine particulate matter and protect health from both primary and secondary pollutants (U.S. EPA, 2017, 2009c,b, 2016).

cost of the policy, $3.2 billion in 2020. In terms of infant lives saved, the ECA had about one-sixth of the effect of the initial 1970 CAA NAAQS (Chay and Greenstone, 2003), one-fifth of the effect of requiring scrubbers at power plants in Germany (Luechinger, 2014), and six times the effect of having cheating diesel emissions (Alexander and Schwandt, 2022). Despite these substantial benefits to human health, the ex-post impact on air quality was weaker than the regulator's ex-ante expectation. Only about 55 percent of the predicted fine particulate matter abatement was realized under the ECA policy.

To better understand why policymakers' expectations were not fully realized, we provide evidence consistent with three types of behavioral responses that altered the policy's effectiveness and were not taken into account in ex-ante predictions. First, we provide evidence indicating that ships altered their routes to avoid using the costly low-sulfur fuel required in the ECA. Second, we provide evidence consistent with "regulatory rebound" in relation to the National Ambient Air Quality Standards (NAAQS) of the Clean Air Act. We find that air quality improvements from the ECA were more muted in counties far from the regulatory threshold for non-compliance with NAAQS, and thus where the risk of crossing the threshold to face regulatory penalties under NAAQS was low. This evidence is consistent with the hypothesis that additional on-land emissions offset some of the decline in emissions from maritime ships. Finally, in addition to difference in the realized air quality improvements, we found gaps in the human health benefits. We estimated that the ECA policy increased time spent outdoors and visits to national park sites, activities which increased individuals' exposure to air pollution. Each of these three types of behavioral responses affected the realized pollution and health benefits of the policy, yet were not incorporated into ex-ante models.

This paper makes several important contributions to the literature. First, we estimate the impact of maritime fuel emissions on air quality and human health. The ECA policy was the first major US environmental regulation of the maritime shipping industry and we are not aware of prior work that evaluates its success. The existing literature on the impacts of ECAs and maritime fuel emissions has relied exclusively on ex-ante prediction approaches or has been conducted in other settings.[3] Predictive models do not take into account compliance or the potential behavioral responses of the regulated industry, other sources, or individuals in response to the policy. Our

[3]See Corbett et al. (2007); Winebrake et al. (2009); U.S. EPA (2009c); Sofiev et al. (2018); Liu et al. (2016); Viana et al. (2020); Zhu and Wang (2021); Lindgren (2021).

ex-post evaluation finds meaningful improvements in air quality from maritime fuel regulation; yet, the improvements are more muted than the predictions from this existing work. We provide a significant advance over the prior literature by documenting that adaptation post-policy critically influences policy effectiveness.

Second, in addition to evaluating the success of the policy, we evaluate the accuracy of the regulator's ex-ante policy analysis. Policy evaluations typically estimate net benefits while ignoring the extent to which those benefits achieved the stated objectives of the policy. Our findings of discrepancies between the policy target and achievement are connected to an existing literature that documents behaviors that diminish the effectiveness of regulation (Becker and Henderson, 2000; Auffhammer and Kellogg, 2011; Fowlie et al., 2016; Zou, 2021). Moreover, our results link shortcomings in the regulator's ex-ante analysis to specific behavioral reactions, including responses by ship operators, other industry, and individuals, and are useful to the design of future policy. These features suggest additions to models to better predict policy effects as well as amendments to policy to improve future regulation of this sector (Duflo, 2017).

Third, this paper contributes to a small but growing literature that incorporates atmospheric aerosol transport models into economics research. Defining where and to what extent the ECA policy affected air pollution for the on-land population is a first-order challenge in this setting. A common approach in economics defines exposure based on distance to the pollution source, but the mobile nature of ship pollution makes this approach difficult. Instead of the "distance" method, we use atmospheric aerosol transport model output as a scientifically grounded prior for pre- and post-policy exposure to pollution.[4] We show that the distance method's failure to account for the complexity of atmospheric interactions can meaningfully reduce precision. Further, our use of transport model output is a new instrument for policy-induced changes in air quality.[5]

The fourth contribution of this work is measuring the infant health effects of transportation emissions in a new setting: at-sea maritime emissions. Infant health has been shown to be sensitive to air pollution and has implications for many later life outcomes, including earnings, cognitive

[4]Some economic research has used atmospheric transport models in other ways. For example, researchers may take estimates of facility-level emission changes driven by regulation and use atmospheric dispersal models to determine impacts of point-source regulation on nearby areas without comparing the model output with in situ observations. For example, Hernandez-Cortes and Meng (2021) analyze resulting changes from cap-and-trade on nearby "environmental justice" gaps, and Sanders and Barreca (2021) analyze the effect of the acid rain program on nearby crop yields.

[5]Since we interpret the transport model output as the policymakers' planned air quality change, this instrument mirrors the method in Baum-Snow (2007).

development, IQ, educational attainment, and welfare take-up (Figlio et al., 2014; Black et al., 2007; Oreopoulos et al., 2008). Prior studies in economics have established a link between infant health and air pollution exposure (Currie and Neidell, 2005; Currie et al., 2009; Arceo et al., 2016), road-vehicle traffic and gasoline content regulations (Currie and Walker, 2011; Knittel et al., 2016; Marcus, 2017; Alexander and Schwandt, 2022), and alternative sources of transportation pollution, such as jets (Schlenker and Walker, 2015). Within this literature, no paper examines a link between maritime fuel content regulation and infant health, even though emissions from shipping fuel have a high concentration of toxic sulfides and comprise a large portion of coastal air pollution. While some work has focused on health effects of port emissions (Moretti and Neidell, 2011; Gillingham and Huang, 2021), we expand our focus to study all at-sea emissions as well, and our results are not driven exclusively by emissions in the vicinity of ports.

Finally, we document how the demographic composition of populations exposed to maritime emissions is distinct from other pollution contexts. The environmental justice literature has documented higher exposure to pollution among disadvantaged populations for many land-based pollution sources.[6] Unlike exposure to stationary pollution sources, we show the proportion of black individuals is smaller for higher levels of maritime emissions. If there are heterogeneous effects of pollution exposure, perhaps due to differences in underlying health conditions, avoidance, or access to care, the realized health effects of maritime emission regulation may be affected by the underlying demographic characteristics of the exposed population. The combination of a demographically distinct exposed population and the unique mixture of pollutants released from ship exhaust makes this an unexamined context in which to explore the impact of maritime fuel regulation, not only on pollution, but also on health.

1 Policy Background

The ECA regulation requires ships to reduce their emissions of air pollutants, primarily sulfur oxides. Figure 1 plots a summary of the policies. Prior to July 2009, the only relevant standard was the IMO global standard, which allowed ships to emit up to 4.5 percent sulfur oxides by mass

[6] Some examples are Superfund sites, hazardous waste sites, landfills, and large polluters from the Toxic Release Inventory (Currie, 2011; Gamper-Rabindran and Timmins, 2011; Banzhaf et al., 2019). Tessum et al. (2021) provide an overview.

(m/m) at any location. This global standard was reduced slightly to 3.5 percent in January 2012.

While the global standard applies to any location, stricter standards can be set near coastlines. In July 2009, California enacted a state standard that allowed at most 1.0 percent sulfur oxides by mass (m/m) in ship emissions within 24 nautical miles of the California coastline. Due to California's limited jurisdiction, however, many ships responded to this restriction by altering their routes to travel just outside the California ECA in order to minimize use of the expensive low-sulfur fuel (Klotz and Berazneva, 2022; Moore et al., 2018).[7] Thus, there was limited scope for the California ECA to improve air quality.

The most significant policy change occurred in August 2012, when the full North American ECA took effect. The North American ECA required low sulfur fuel of up to 1.0 percent sulfur oxides. The regulation applied within the exclusive economic zone of the participating countries, the United States and Canada, depicted in Figure 2. This jurisdiction extended for 200 nautical miles from the coast except in parts of southern California, Texas, and Florida where it was equidistant from non-participating neighboring countries, Mexico, the Bahamas, and Cuba. While 200 nautical miles jurisdiction limited the scope for avoiding the use of low-sulfur fuel, in the areas with reduced jurisdiction, the high cost of low-sulfur fuel may have created sufficient incentives for ship relocation to avoid regulation, similar to the response observed when ships avoided California's narrow ECA.

In subsequent years, the fuel content restrictions were tightened. In January 2014, California made its state standard more stringent: it allowed up to only 0.1 percent sulfur oxides. In January 2015, the full North American ECA also reduced the allowance to 0.1 percent sulfur oxides.[8] The tightening of these standards may have led to a growth in the effect of the policy over time.

In addition to requirements for the use of lower sulfur fuel near coastlines, the ECA regulation tightened standards for engine emissions of nitrogen oxides. These additional standards applied

[7] Using detailed ship location transponder data, Klotz and Berazneva (2022) find a sharp reduction in distance traveled, speed, and fuel consumption within California's ECA, along with an even larger increase in fuel use just outside the ECA due to ships traveling greater distances, and in some cases higher speeds, to avoid the California ECA. Similarly, we failed to find an effect of California's ECA using 2007-2011 data from California only. We modified equation 1 such that the post-policy indicator is equal to one after California's ECA is in place, July 1, 2009. Results are available upon request. California altered its emission control area in December 2011 by extending a portion of the boundary to include the area around the Channel Islands in an effort to encourage ship traffic to return closer to shore.

[8] As with other environmental standards, the pattern of California preceding federal environmental standards with strict state standards arguably motivated coordinated action from industry groups and the federal government.

to only a small subset of ship traffic: new US-flagged ships delivered after the policy came into effect. Since this aspect coincides with the sulfur oxide regulation, we cannot separately estimate its contribution; however, we expect this to have a very small effect in the years immediately after implementation because only a small percentage of total ships were subject to this requirement.

The new standards applied to all commercial ships and distributors of marine fuel. To reduce emissions, ships could use compliant fuel or an approved equivalent method, such as a scrubber. Although compliant fuel was more costly than typical bunker fuel, ships were already equipped with multiple fuel tanks and could easily switch to a tank with compliant fuel as they approached the regulated area. Scrubber installation required investment in new equipment and was uncommon except among cruise and passenger ships (Hellenic Shipping News, 2014). The US Coast Guard (USCG) was responsible for enforcement and ensured compliance through scheduled and unscheduled inspections.[9] Vessel operators had to provide documentation of fuel purchase and delivery, fuel samples, written fuel oil changeover procedures, and a fuel oil changeover log book that records the volume of compliant fuel in each tank as well as the date, time and position of the ship when any fuel oil changeover operation was completed. Further, the sale of non-compliant fuel was outlawed. Violations were governed by the provisions in the Act to Prevent Pollution from Ships. Non-compliance was penalized with fees of up to $25,000 for each violation, and each day of continuing violation could constitute a separate offense. In cases where an incoming ship could establish that compliant fuel was not available, it could appeal for an exemption from penalties.

2 Data

Our analysis combines data on EPA air quality predictive models, observed air pollution, infant health, mortality, weather, county characteristics, and outdoor activities, which we describe below.

Air Quality Model and Data. For our measure of intensity of exposure to the policy, we employ the EPA's Community Multiscale Air Quality Modeling System. Our main treatment variable is

[9] The USCG could check for ECA compliance during normally scheduled port state control exams, domestic vessel inspection, and vessel safety examinations. Vessel operators were required to demonstrate compliance to USCG port state control examiners, marine inspectors, and boarding officers who attend vessels for a variety of purposes both in port and at sea.

the predicted reduction in PM2.5 as a result of the ECA regulation. The EPA developed these predictions as a component of their proposal to justify the ECA policy (U.S. EPA, 2009c). We obtained the output of the CMAQ ECA analysis in 12km resolution raster grids for (i) 2020 annual mean PM2.5 concentration under business as usual and (ii) 2020 annual mean PM2.5 concentration under the ECA regulation. Our independent variable of interest, CMAQ change, is the value at the county population-weighted centroid of (i) minus the county average of (ii) and is shown in Figure 3. The CMAQ predictions are based on 2002 ship traffic and fleet characteristics. Traffic is scaled to approximate 2020 ship traffic levels but is not adjusted for behavioral adaptations in shipping activity as a result of the ECA regulation.

We limit our sample to the counties shown in Figure A1 whose centroids are within 200km of heavy ship traffic in 2010, which is defined as the top 5th percentile of raster grid cells. Counties more than 200km from heavy ship traffic are less suitable as controls, but we show in robustness checks that our results are not sensitive to this sample selection criterion.

Our main air quality outcome is fine particulate matter (PM2.5). Although sulfur dioxide (SO2) emissions were regulated at sea, sulfur dioxide does not last in the atmosphere for long periods, nor does it travel significant distances. We focus on over-land secondary PM2.5. PM2.5 is both a direct and secondary pollutant of ship exhaust, and over-land secondary PM2.5 was the criterion pollutant targeted by the ECA fuel content regulation (U.S. EPA, 2009c).[10]

Our air quality data comes from the United States Environmental Protection Agency Air Quality System (AQS) database. We average monitors within-county to construct county-month mean air quality. Observations are missing if a monitor is scheduled to be down for maintenance, if the collection does not meet the data quality standards, or if a new monitor location is introduced mid-sample. To ensure against bias arising from these events, we only use monitors that were observed at least once per year from 2008 to 2016.[11] More details are provided in Appendix Section 9.1.1. In robustness exercises, we also use data on counties' air quality performance relative to the National Ambient Air Quality Standards (NAAQS), described in detail in Appendix Section 9.1.2.

[10]Nitrogen oxide and its derivative, ozone, were separate components of the ECA regulation. We do not include these pollutants because the regulation targeted them with a slowly phased-in engine requirement, and we do not expect to capture the effects of this component with our research design.

[11] Figure A1 shows counties with balanced and unbalanced monitors. In Table A5, we show that our results are robust to relaxing our balanced monitor requirement.

Health Data. Infant health data comes from the National Center for Health Statistics Vital Statistics Natality records from 2007 to 2017. Counties with few births are excluded for anonymity. The sample is restricted to singleton births, hospital births in the continental US, mothers between ages 18 and 45, and births with non-missing birth date, birth weight, and gestation.

Birth weight is measured in grams, with newborns under 2,500g classified as low birth weight. Gestation is measured in weeks; births before 37 weeks are classified as preterm births. Mother-infant variables included as controls are indicators for mother's years of education $\{< 12, = 12, 13-15\}$, mother's race {black}, Hispanic, mother's age $\{19-24, 25-34, > 35\}$, two previous live births, three or more previous live births, and cigarette use during pregnancy. For each control variable, an indicator is included for missing observations.

We map births to month of conception based on the reported gestational age and month of birth. We collapse birth observations to county-month cells for computational efficiency. The incidence of low birth weight and preterm births are measured per 1,000 births. Observations are weighted by the number of conceptions unless otherwise noted.

We supplement the birth outcome data with county-level mortality data from the National Center for Health Statistics Vital Statistics Mortality records from 2007 to 2016. We calculate the death rates per 1,000 population using age-specific county population measures from the Surveillance, Epidemiology, and End Results Program (SEER) data. Infant mortality is measured as the number of deaths among children under age one per 1,000 births.

Weather and Other Data. Weather is an important control because air quality is highly dependent on meteorological conditions that transport and disperse air pollution. These meteorological conditions also directly affect infant health (Barreca and Schaller, 2020). We use the PRISM Daily Weather Data for the Contiguous United States (Schlenker, 2020). This data features a balanced panel of weather station records from 1950-2018 that are combined to daily 2.5 by 2.5 mile grids of minimum temperature, maximum temperature, and total precipitation. 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 (see Appendix for more detail). Our results are also robust to controlling for more flexible weather bins.

In addition, we adjust for local economic conditions using county-month unemployment rate from the Bureau of Labor Statistics Local Area Unemployment Statistics (LAUS). Finally, we measure outdoor activity in order to observe whether individuals exhibit behavioral changes in response to changing air quality. Data from Recreation.gov maintains information on millions of visitors to federal parks, and data from the American Time Use Survey (ATUS) from 2008 to 2016 measures time spent outdoors based on a detailed time diary of all activities over a 24-hour period. We describe these sources in more detail in Appendix Section 9.1.4.

Table 1 provides summary statistics for counties in our sample. Statistics are weighted by the number of conceptions in the county-month. Column (1) reports means for all counties within 200km of heavy ship traffic, and column (2) restricts the sample of counties to only those with a balanced sample of air quality monitors. The samples appear very similar across all outcomes and control variables. For this sample, the average level of fine particulate matter is about 9 $\mu g m^{-3}$, about 6 percent of births are classified as low birth weight, and average birth weight is about 3,300 grams.

3 Impacts of ECA on Air Pollution and Health

3.1 Methods

To estimate the causal effect of the ECA regulation on air quality and health outcomes, we exploit variation from the policy timing and intensity across locations. The intuition of our approach is that we compare changes in outcomes in counties that were highly exposed to pollution from ship exhaust relative to changes in outcomes in counties that were less exposed to pollution from ship exhaust, before and after policy adoption.

While distance is commonly used to proxy for intensity of exposure to pollution in other contexts, other factors influence exposure. Exposure to emissions from ship exhaust is a combination of ship traffic, fuel content, distance, atmospheric interactions, and meteorological factors that disperse emissions. Including interactions of ship traffic, distance, and weather to proxy for exposure presents several concerns for estimation. Instead, we employ the predictions of an atmospheric aerosol transport model to combine these components into an exposure index. This approach provides several advantages.

First, ex-post observations of ship traffic and emissions are likely endogenous to the policy. Although ex-post observations of ship traffic/emissions can yield quasi-exogenous short-run variation in air pollution, which can be used to estimate effects of air pollution (as in Knittel et al. (2016) and Moretti and Neidell (2011)), ex-ante observations are more appropriate for determining policy effectiveness, for a few reasons. First, it is plausible that ship traffic falls as a result of the regulation if it is no longer profitable to deliver to US ports or if ships alter routes to avoid traversing the regulated areas. If the econometrician used ex-post traffic as a metric of exposure to the policy, they would fail to attribute pollution change from lower traffic to the policy. Second, ex-post ship traffic reflects economic conditions that influence other sources of air pollution as well as infant health. The econometrician risks overstating the changes from the policy if the measure of exposure is correlated with other changes in air pollution.

A second primary concern with using interactions of ship traffic, distance, and weather is bias from assuming an incorrect functional form of their interaction. For example, assuming a linear relationship between destination air quality and distance to source overstates the contribution of the source at distances beyond its average dispersion range. While the econometric methods exist to fit the data and determine an appropriate model specification, this exercise is cumbersome because transportation of air pollution and the creation of secondary pollution depends on many combinations of atmospheric conditions that vary by source location, destination location, and time, among others. By contrast, output from aerosol transport models incorporates these various factors *a priori*.

For these reasons, we employ output from the EPA's Community Multiscale Air Quality Modeling System as our measure of intensity of exposure to the policy. The EPA developed these predictions as a component of their proposal to justify the ECA policy (U.S. EPA, 2009c). Our main treatment variable is the predicted reduction in PM2.5 as a result of the ECA regulation, and represents the policymakers' ex-ante expectations of the policy's effects on air quality.

We start by estimating the reduced form effect of exposure to the policy, as measured by the CMAQ prediction, on each of our outcomes of interest. Denote county i in year-month ym, where m indicates the calendar month (January-December) and y indicates the year (2008-2016). The outcomes of interest Y_{iym} are the mean air pollution, PM2.5, and health measures. The main health outcomes are rate of low birth weight and preterm birth for births conceived in county i in year-month ym along with overall mortality rate and infant mortality rate for county i in year-month ym. The exposure variable is $CMAQ_i$, which is the CMAQ prediction of the reduction in PM2.5 due to the ECA.

We estimate the overall reduced-form effect of the policy in the post-period with the following difference-in-difference specification:

$$Y_{imy} = \beta CMAQ_i \times postECA_{my} + \delta X_{imy} + \tau_{ry} + \alpha_{is} + \epsilon_{imy} \tag{1}$$

where $postECA_{my}$ is an indicator equal to one after the ECA policy came into effect in August 2012, τ_{ry} are region-by-year fixed effects (i.e., Gulf Coast 2008, Gulf Coast 2009, ... Gulf Coast 2016), α_{is} are county-by-season fixed effects (i.e., Marin County Spring, Marin County Summer, ...), and X_{imy} are additional controls for county-month-year weather and unemployment rate. Including region-by-year fixed effects allows for differential trends or shocks by region and makes counterfactual comparisons only within the same region. County-by-season fixed effects allow 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. However, we show our results are robust to a variety of alternate sets of controls and fixed effects. In the baseline estimation of (1), county-month-year observations are weighted by the number of conceptions in county i in year-month ym. For mortality outcomes, we weight observations by age-specific county population. When estimating (1) with infant health outcomes, mother-child covariates are included and weather controls are included by trimester. For example, X_{imy} includes the max temperature in the first trimester, second trimester, and third trimester for conceptions in each year, month, and county. Robust standard errors are clustered by county.[12]

For our estimates to measure the effect of the policy, we must assume that there are no omitted time-varying, county-specific features correlated with the ECA timing and exposure that also affect our outcomes of interest. This assumption would be violated if, for example, another envi-

[12] Although each county centroid is within a unique CMAQ grid cell, we repeat the main results with standard errors clustered at the state level to address potential dependence. Clustering at the state level is a more conservative approach than clustering at the grid point as it also allows for spatial correlation within a state (Bester et al., 2016).

ronmental regulation came into effect at the same time and its intensity was correlated with ship pollution exposure. These concerns are mitigated by the inclusion of controls for arbitrary region-year shocks, as well as arbitrary county-season seasonality. In robustness exercises, we show the results are robust to including state-year fixed effects and also to controlling for county compliance with environmental regulations. Any violation of the identifying assumption would need to follow the same timing and county specific-intensity as the ECA regulation for the violation to affect our outcomes.

The absence of such violations implies that outcomes do not trend differently between counties with higher and lower anticipated CMAQ pollution reductions in a world without the policy. To support this assumption, we estimate event-study specifications and show that an additional unit of predicted CMAQ pollution reduction does not affect the trend in air quality in the years before the ECA implementation. We also show that maternal demographic characteristics and other placebo outcomes are not changing systematically with the policy variation.

A distinct advantage of using the CMAQ reductions as the "treatment" variable is the ability to compare the realized changes in air quality to the intended changes in air quality. Even if the CMAQ model is imperfect, the CMAQ-predicted reductions represent policymakers' plans. If each unit of planned PM2.5 reductions yielded one unit of actual reductions, the measured effect, β, would equal -1. Consequentially, we interpret estimates for PM2.5 that deviate from -1 as evidence that the intentions were not fully realized.

3.2 Results: Air Pollution

First, we show the direct effect of maritime fuel regulation on air pollution. Figure 4 plots the event-study for fine particulate matter. The figure shows that air pollution was not trending differently in counties with different levels of CMAQ-predicted pollution improvements prior to the regulation. We fail to reject the null that each of the pre-policy coefficients is statistically different than zero. In the post-regulation period, there is a decline in PM2.5 among counties with greater exposure to ship traffic. We note that 2012 is a partially treated year because the policy was implemented in August 2012. We reject the null that the post-policy coefficients are jointly equal to zero.

Column 1 in Table 2 reports the difference-in-difference coefficient from estimation of equation (1). Consistent with Figure 4, there is a significant reduction in fine particulate matter after the ECA policy was implemented. One additional unit (μgm^{-3}) of predicted reduction in PM2.5 led to a 0.55 unit (μgm^{-3}) fall in PM2.5 after the policy. Relative to the average level of fine particulate matter, this represents a 6 percent decline. To measure the overall effect on fine particulate matter, we scale the county-level CMAQ-predicted improvement by our estimated coefficient (0.55) and take a population-weighted average across all counties within 200km of heavy ship traffic. Figure A2 shows the scaled county-level fine particulate matter improvements. We calculate that fine particulate matter decreased by about 0.37 units on average, or about 4.0 percent relative to the mean of 9.25 units.

While this result is statistically significant and economically meaningful, the fine particulate matter point estimates indicate that air pollution fell by roughly 55 percent of the amount policymakers intended when they designed and implemented the policy. The coefficient is statistically significantly different from one (p < 0.001).[13] Therefore, we can reject the hypothesis that the decline in fine particulate matter forecasted by the CMAQ model was realized.

3.3 Results: Health

Even though the forecasted effect of the ECA on air pollution was not fully realized, the policy still led to meaningful improvements in health. We start by looking at the reduced-form policy impacts on health. Figure 5 shows the event-study figures using the low infant birth weight rate and preterm birth rate as the outcomes. Here, the omitted period is conceptions in 2010 because conceptions in 2011 and 2012 were partially exposed to the policy during gestation. Conceptions in 2013 and after were exposed to the policy for the entire duration of the pregnancy.

The patterns in the infant health measures are consistent with the trends in air pollution in Figure 4. In support of the parallel trends assumption, we found CMAQ did not predict changes in infant health relative to the omitted period in the years before the regulation for each of the infant health outcomes. Panels (a) and (b) indicate that the ECA regulation led to significant improvements in the rates of low birth weight and preterm birth. As expected, the effect is somewhat muted

[13] A more conservative approach discounts the reference point to align the CMAQ projections, which are the effect in 2020, with the coefficient, which is the average effect for roughly 2013-2016. When we make this adjustment using the annual growth rate of 3.4% from (U.S. EPA, 2009c), we conclude 67% of the forecasted improvement was realized and still reject the null that the coefficient equals the conservative reference point of 0.83 ($p < 0.001$).

for conceptions in 2011 and 2012, because these observations were only partially exposed during gestation to the ECA policy. For fully treated cohorts, 2013 and after, the coefficients indicate a sizable improvement in low birth weight and preterm birth rates.

Columns (1) and (2) in Panel A of Table 3 show the corresponding overall difference-in-difference regression results from estimation of the reduced-form equation (1) for each of the infant health measures. Additional infant health measures are reported in Appendix Table A1. The effect on all outcomes is statistically significant. For low birth weight, one additional unit of CMAQ-predicted PM2.5 reduction lowered the rate by 1.3 births per 1,000 after the introduction of the ECA, a 2 percent improvement relative to a baseline of 61 low weight births per 1,000 births. Using the mean CMAQ-predicted change, 0.76, this coefficient suggests that the number of low birth weight infants declined by 1 per 1,000 births, or 1.7 percent relative to the mean. One unit of CMAQ-predicted change led to 2.1 fewer preterm births per 1,000, or 2 percent. For the mean CMAQ-predicted change, 0.76, the coefficient suggests preterm births declined by 1.6 per 1,000, or 1.7 percent.

We also report results for mortality in Columns (3)-(4). Column (3) reports the overall effect on the all-age death rate per 1,000 population. A one-unit predicted change in fine particulate matter measured by CMAQ leads to a statistically significant reduction of 0.006, or about 1 percent relative to the mean, after the ECA is adopted. For the average CMAQ-predicted change, 0.76, this is a change of about 0.7 percent. Column (4) reports the effect on infant mortality (under age one). A one-unit predicted change in PM2.5 based on CMAQ leads to a statistically significant 0.24 percentage point, or 3.7 percent, reduction in the rate of infant mortality (2.8 percent for the mean CMAQ prediction). Event study results confirm these results in Figure 6. For both overall mortality and infant mortality, there is little evidence of a pre-trend in years prior to policy adoption, but there is a decline in mortality in areas with heavy ship traffic following adoption of the ECA.

3.4 Additional Results and Robustness Checks

Alongside the main results, we provide a detailed analysis of additional results and robustness checks in the Appendix 9.2. We briefly summarize those findings here.

We first examine a number of placebo outcomes that support the validity of our results. A potential concern with our estimation of the ECA's effects is that the introduction of the ECA could have been correlated with other changes, such as shifts in demographic characteristics or local economic activity, that would have led to improvements in the outcomes in the absence of the policy. 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 of any such patterns driving the results. As shown in Table 2 column (2) and Figure A6, we failed to find that an index of demographic characteristics is 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. Furthermore, Table A4 shows there is no significant relationship between the policy variation and the following additional outcomes: pollutant emissions from power plants, fine particulate matter emissions from all sources, the number of other Toxic Release Inventory (TRI) pollution sources, 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.

Second, we examine a number of additional health outcomes and find a consistent pattern of improvements. First, we find consistent significant improvements in alternative measures of infant health, average birth weight and gestation. To examine the ECA's effects on the distribution of birth weight, we measure the effect of the policy on bins of birth weight. 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, where improvements are quite beneficial. Similarly,the elderly also tend to be particularly sensitive to air pollution, we find that the ECA led to statistically significant declines in mortality for individuals age 75-84 and age 85 and over of 0.8 and 1.4 percent respectively. As a validly check, we failed to find evidence of pre-trends in years prior to the policy for elderly mortality.

Third, we look for air quality improvements with respect to other pollutants. As an alternative measure of air quality, we find a significant improvement in the overall Air Quality Index. We also find evidence that sulfur dioxide and nitrogen oxides could have played a role in the air quality improvement form the ECA; however, we interpret these results with caution because the CMAQ model predictions for PM2.5 will not accurately capture the spatial impacts on other pollutants and few monitors for other pollutants are balanced over a long panel.

Fourth, we consider whether 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. Results reported in Table A7 show meaningful improvements in precision when CMAQ is used to measure exposure to the ECA policy.

Finally, we perform a number of robustness and sensitivity checks in Table A5 and Table A6. To start, we demonstrate no consequential change in precision when we cluster the standard errors at the state-level. We then show that our results are robust to alternative choices for inclusion in the sample in lieu of limiting the sample to counties whose centroids are within 200km of heavy ship traffic. Similarly, we relax the balanced panel requirement for air quality monitors and include more flexible weather controls without meaningful change in the conclusions. We also use an alternate measure of intensity of treatment that is based on the CMAQ prediction of total emissions from maritime shipping.

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 another exercise, we build up to our preferred specification to show our results are robust to a variety of alternate sets of fixed effects. We also report consistent results for a binary difference-in-difference model.

Last, we address the possibility that our results could reflect other pollution abatement policies during our sample period. First, we repeat our analysis dropping port counties and find our results are not driven by port counties alone. This exercises shows our results are not driven by any port-specific policy changes that may have been adopted during our sample period. Second, we show 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 fuel standards, we show that the reduced form effects of the ECA on main outcomes are robust to including NO2 as a control variable. Finally, we failed to find that the tightening of the fuel content standard in 2015 had an additional impact on air quality. This is not surprising, as the 2015 fuel standard tightening was a relatively small change and many ships were already using compliant fuel.

4 Comparison to Other Sources and Incidence

This section compares the marginal benefits of abatement from maritime shipping to those of other settings. Although many characteristics of maritime pollution differ from other sources of pollution, we focus on quantifying differences in exposure and effects across demographic groups.

4.1 Exposure Across Demographic Groups

To understand how the population exposed to maritime traffic pollution compares the population exposed to land based sources, we begin with descriptive evidence. Despite existing evidence of disproportionate pollutant exposure for disadvantaged groups from land-based pollution sources, prior work has not measured the exposure gap for a source that is mobile and at-sea. To maintain similarity in the setting, we use the distance to the nearest port as the representative of a comparable stationary, land-based source of pollution. We highlight the differences in the demographics of the population exposed to maritime emissions, measured by the CMAQ model versus ports, measured by distance to ports. Details of the analysis summarized here are in Appendix.

Figure 8 presents the cumulative distribution function of exposure defined by distance to port, panel (a), and ship emissions, panel (b), for each demographic subgroup. Panel (a) shows that non-Hispanic blacks and other non-white individuals are roughly equally likely to live near a port and all non-whites are more likely to live near a port than whites. This pattern of disproportionate exposure for non-whites is consistent with the large environmental justice literature looking at stationary land-based pollution sources (Tessum et al., 2021). However, the exposure pattern is different with respect to maritime emissions in panel (b). Panel (b) shows that black and white individuals have almost identical exposure to maritime emissions even though black individuals live closer to ports than white individuals on average. Further, unlike with land based sources, there are differences in exposure to maritime emissions across non-white subgroups. Hispanics and non-Hispanic other race individuals are more exposed to maritime shipping emissions than black individuals.[14]

Given that the exposed population is different from most land-based pollution sources, the health effects of this policy could differ from other reductions in air pollution. This is likely if 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.

4.2 Methods

In this section, we use the ECA policy variation to instrument for fine particulate matter to estimate the impact of reductions in fine particulate matter in this setting on health. This estimate allows for a comparison of the marginal health benefits of pollution reductions from maritime shipping with those of other sources.

The first stage specification is provided in equation (1) and the second stage is shown below:

$$health_{imy} = \gamma \widehat{PM2.5}_{imy} + \delta X_{imy} + \tau_{ry} + \alpha_{is} + \epsilon_{imy} \tag{2}$$

Using $CMAQ_i * postECA_{my}$ as an instrument for PM2.5 exploits the fact that some portion of the PM2.5 changes in each county occurred simply because the county was more exposed to the baseline dispersion of emissions from maritime traffic while excluding the changes in PM2.5 that occurred because of adjustments to the pattern ship traffic, amount of other emissions, etc. The exclusion restriction requires that the ECA policy implementation affects health only through its effect on pollution. Although we cannot formally test this assumption, we demonstrate that the ECA policy was uncorrelated with observable demographic and economic changes in Tables 2 and A4.

While we focus on fine particulate matter as our primary pollutant of interest, we cannot rule out that the ECA also resulted in modest improvements in sulfur and nitrogen oxides. To that extent, our estimates capture the effect of a mix of fine particulate matter and these other pollutants.

[14] In the Appendix, we observe consistent patterns when we show the correlation between race/ethnicity and our two measures of exposure Figure A8.

This problem plagues most estimates of the health impacts of air pollution because air pollutants are inherently correlated. Nevertheless, this exercise helps put our estimates into context with the existing literature.

4.3 Results: 2SLS and Comparison to Other Sources

Panel B in Table 3 first reports the two-stage least squares results from estimation of equation (2), while Panel C reports the ordinary least squares results. Instrumenting for fine particulate matter with our policy variation, we find that a one-unit increase in fine particulate matter leads to 2.7 additional low birth weight infants per 1,000, or a 4.4 percent increase relative to the mean. Columns (3) and (4) of Panel B show that a one-unit increase in fine particulate matter leads to a 1.7 percent increase in overall mortality and an 6.5 percent increase in infant mortality.

We compare the magnitude of our estimates to the literature in Table 4, following Alexander and Schwandt (2022), who consider the effect of a 10 percent pollution increase. Our results suggest that a 10 percent increase in pollution would increase low birth weight by 4.1 percent and infant mortality by 6.0 percent. The magnitude of the effect for infant mortality is in line with the recent literature focusing on fine particulate matter, while our estimate for low birth weight is slightly smaller. This could be due to the unique bundle of pollutants impacted by the regulation or to the differences in the demographics of the population most exposed to the regulation.

4.4 Results: Incidence across Demographic Groups

As documented in section 4.1, the demographic composition of the population most exposed to ship traffic, as measured by the CMAQ model, differs from many land-based pollution sources, which may drive differences in the health effects that we observe relative to the previous literature. Next, we explore the degree to which the ECA has a heterogenous effect on the exposed population.

We estimate two-stage least squares from equation (2) on subsamples of mothers by demographic characteristics at the individual level. Table A8 shows the results by race/ethnicity, education, age, and marital status. Column (1) reports the baseline estimates for the full sample at the individual level. The magnitude of the effect is similar to the main results for low birth weight in Panel B of Table 3, which both show about a 0.2 percentage point, or 4 percent, increase in low birth weight for a one-unit increase in fine particulate matter. Columns (2) - (5) show results for non-Hispanic white, non-Hispanic black, non-Hispanic other, and Hispanic mothers. The magnitude of the effects are largest for non-Hispanic white mothers and non-Hispanic other race mothers. The estimated effects show that a one-unit increase in fine particulate matter leads to a 6.6 percent and 12.6 percent increase in low birth weight for non-Hispanic white and non-Hispanic other race mothers, respectively. The results for mothers with high education in column (6) are similar in magnitude to the overall effects, suggesting that heterogeneity is driven less by education level. Column (7) reports results for married mothers only. The magnitude of the effect is only slightly smaller for married mothers, still about 4 percent from the mean. Finally, columns (8) to (10) report results for mothers age 19-24, 25-34, and over 35. The main results appear to be driven by mothers over 25.

5 Behavioral Responses

While the effect of the ECA regulation led to a statistically significant improvement in fine particulate matter and health, the estimated effect was less than anticipated. One explanation is that the CMAQ model did not take into account behavioral changes. In this section, we explore whether there is evidence of any behavioral change along three dimensions in response to the ECA: ships, other emissions, and individuals.

5.1 Shipping Behavior

We hypothesize that ships most likely exhibited behavioral responses that diminished the effectiveness of the policy in coastal areas where the cost of avoiding the ECA is lowest. As shown in Figure 2, because Mexico did not participate, southern parts of California, Florida, and Texas were less than 200 nautical miles from the exterior of the ECA. In these areas, it was less costly to travel to exit the ECA and avoid using costly low-sulfur fuel, and the use of higher-sulfur fuel outside the ECA was nearer to coastal populations. In addition, Klotz and Berazneva (2022) provide evidence that a narrow 24 nautical mile boundary led to substantial behavioral response among ships. Therefore, we hypothesize that areas fully exposed to the ECA, at least 200 nautical miles from the exterior of the ECA boundary, had larger impacts on air quality than areas with only partial exposure to the ECA. We estimate,

$$
\begin{aligned}
Y_{imy} = &\beta_1 full_i \times CMAQ_i \times postECA_{my} + \beta_2 partial_i \times CMAQ_i \times postECA_{my} \\
&+ \delta X_{imy} + \tau_{ry} + \alpha_{is} + \epsilon_{imy}
\end{aligned} \tag{3}
$$

where $full_i$ equals one for counties exposed to the full 200 nautical mile ECA and $partial_i$ equals one for counties less than 200 nautical miles from the exterior of the ECA. Other variables are defined analogously to equation (1). Because we are interested in the spatial distribution of pollution reductions in this part of the analysis, rather than health effects, we do not weight by population. Standard errors are clustered by county.

We estimate an event study specification to test this hypothesis in Figure 7. Neither panel shows evidence of a pre-trend in fine particulate matter prior to the ECA implementation. In panel A, for the areas exposed to the full ECA, there is a clear and statistically significant decline in fine particulate matter after policy adoption. However, in panel B, areas with only partial exposure to the ECA show a somewhat noisier and more muted effect of the policy, as expected.[15] In terms of magnitude, the post-policy coefficients in panel A are not statistically distinguishable from -1 in each year from 2013 to 2016, suggesting that the anticipated declines in fine particulate matter were realized in areas fully exposed to the 200nm boundary. By contrast, the estimated coefficients in panel B suggest that the ex-post decline in PM2.5 was less than the anticipated decline in areas only partially exposed to the ECA.

Table A9 summarizes these effects. First, column (1) replicates the effect of the ECA in the full sample without population weights. The coefficient is very similar to column (1) of Table 3 and shows the decline in fine particulate matter was a little more than half of the expected decline overall. Column (2) estimates equation (3) for counties partially and fully exposed to the 200nm boundary. A one-unit increase in CMAQ is associated with a 0.86 and 0.49 unit (or 10 percent and 6 percent) decline in fine particulate matter in areas fully and partially exposed to the policy,

[15] In the Appendix we explore the source of the year-to-year variation in the partially exposed counties. Including differential yearly trends in southern California eliminates the year-to-year variation in the post period, 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.

respectively. The decline in fine particulate matter in areas partially exposed to the ECA is statistically significantly smaller than the decline in fine particulate matter among fully exposed areas (p < 0.05). This is consistent with the hypothesis that behavioral response among ships in areas partially exposed to the ECA, where avoiding the ECA was easiest, led to a muted effect of the policy on air pollution. However, in areas fully exposed to the ECA, the CMAQ-predicted reductions in fine particulate matter were statistically indistinguishable from the realized reductions, suggesting the policy was effective in these areas.

5.2 Other Emissions Behavior

Alongside changes in the location of ship emissions, we hypothesize that other polluters could also respond to the implementation of the ECA policy. For example, the regulator was especially concerned that the increased in costs for ECA-compliant fuel would lead to an increase in more highly-polluting land-based transportation. Cost estimates showed such a shift would not be economical U.S. EPA (2009c,a). Instead, we focus on examining another form of behavioral response the regulator did not consider: regulatory interaction with the National Ambient Air Quality Standards (NAAQS). The Clean Air Act requires counties to maintain ambient fine particulate matter concentrations below the NAAQS or else face high regulatory cost, including state implementation plans, pollution monitoring, and new source review. We hypothesize that counties that experience air quality improvements due to ECA have less need to engage in costly efforts to reduce pollution from other sources to maintain compliance with the NAAQS. We refer to this type of response as rebound.

Three features restrict the scope for emissions rebound. First, rebound will only occur where the NAAQS result in costly pollution reduction efforts, such that pollution is below the unconstrained optimum. Second, since non-attainment of the NAAQS results in costly regulation, we expect rebound is unlikely when it might lead a county to enter non-attainment or to remain in non-attainment.[16] Third, counties must also have the ability to change emissions relatively easily.[17]

[16]Importantly, there is uncertainty as to whether a given level of emissions would result in ambient air quality that leads to non-attainment. Sources of uncertainty include outside pollution, weather, legal conditions required for monitor observations to be considered for design values, etc.

[17]Response could take a number of forms: (1) altering the issuance of air quality alert days can impact pollution

We expect that these incentives for a rebound effect vary across counties as a function of their risk of falling into non-attainment status. Counties with the lowest pollution levels have low incentives to rebound if they are not constrained by the NAAQS. Because our sample restricts to counties with consistently monitored air quality recordings over the entire sample period, we exclude counties with the lowest pollution levels that are least likely to be constrained by the NAAQS. Counties with higher pollution levels are more likely to be constrained, but only counties sufficiently far from the threshold might benefit from increasing emissions without risk of entering or remaining in non-attainment. Counties above the threshold may not have the ability to rebound if regulators carefully monitor emissions, such as through new source review. However, it is difficult to define a domain of counties that have a flexible source of emissions and are not so "close" to the non-attainment threshold that rebound becomes too risky. Within our sample, we expect the likelihood for rebound will be lowest near the regulatory threshold. Although we cannot rule out rebound in counties above the regulatory threshold a priori, we note that the restrictions on new emissions in these locations limit their flexibility for rebound. Therefore, we expect the highest potential for rebound for counties sufficiently below the regulatory threshold so as to not risk entering non-attainment.

To examine this pattern, we allow the effect of the ECA to differ along with the distance to the regulatory threshold. We estimate,

$$Y_{imy} = \sum_k \beta_k \mathbb{1}[D_i \in k] \times CMAQ_i \times postECA_{my} + \delta X_{imy} + \tau_{ry} + \alpha_{is} + \epsilon_{imy} \tag{4}$$

where D_i represents county i's pre-policy distance to the regulatory threshold. We define D_i as the county 2012 PM2.5 maximum design value as a percent of the standard.[18] We classify this distance into seven bins of 2012 PM2.5 as a fraction of the standard: less than 60%, 60-70%, 70-80%, 80-90%, 90-100%, 100-110%, and over 110%. Other variables are defined as in equation (1). The estimates are unweighted and standard errors are clustered at the county level.

from cars, (2) changing emissions at a source that has flexibility in production duration/timing, (3) altering strategic behavior on known monitoring days, etc. Unfortunately, we cannot measure each of these potential response channels.

[18] We obtained the EPA records used to determine compliance with the NAAQS. For 2012, the NAAQS required counties' PM2.5 to meet two thresholds: (i) annual mean PM2.5 averaged over three years, $\text{DV}^{\text{1 year}}$, less than $12\mu\text{gm}^{-3}$ and (ii) 98th percentile of daily mean PM2.5 averaged over three years, $\text{DV}^{\text{24 hours}}$, less than $35\mu\text{gm}^{-3}$. We then defined distance to the regulatory threshold, $D_i = 100 * \max\{\frac{\text{DV}^{\text{1 year}}}{12}, \frac{\text{DV}^{\text{24 hours}}}{35}\}$.

Figure A10 reports the results of estimating equation (4). We found a pattern consistent with our hypothesis that increases in on-land emissions offset declines from at-sea emissions: as the county's risk of violating the regulatory threshold decreased, the impact of the ECA was more muted. Although rebound could have occurred elsewhere, this evidence indicates it is concentrated in counties below 80 percent of the regulatory standard. The counties at greatest risk of violating the Clean Air Act threshold, from 90 to 100 percent of the regulatory standard, experienced the greatest declines in PM2.5 as a result of the ECA policy. Although this decline appears larger than what the ECA plausibly delivered, we fail to reject the null that the magnitude is equivalent to -1.

Table A9 column (3) summarizes the differential effects of the ECA for counties further from the Clean Air Act regulatory threshold relative to counties close the regulatory threshold. It reports the estimates of a variation of equation (4) where the counties with pre-policy PM2.5 within 90 to 100 percent of the regulatory threshold are the omitted category. We found that all counties had significantly smaller declines in PM2.5 resulting from the ECA relative to the group that was closest to violating the Clean Air Act. The most significant offsetting effects appear to have occurred in counties below and furthest from the threshold. The results suggest a rebound from other emissions that entirely offset the air quality improvements from the ECA among counties with pre-policy PM2.5 below 80 percent of the regulatory threshold. However, we underscore that other features of the setting could have also diminished the effect of the policy for these counties.

Overall, our evidence is only suggestive of the premise that rebound emissions muted the benefits of the ECA policy. Additional results in the Appendix report consistent yet inconclusive estimates of an increase in reported emissions corresponding to rebound incentives. Unpacking the extent of regulatory interaction between the CAA and other policies is an important area for future research.

5.3 Individual Behavior

Next, we explore whether policy-induced improvements in air quality had a subsequent impact on individuals' behavior. Increased time spent outdoors, for example, could increase individuals' duration of exposure to the now lower level of air pollution. Because ex-ante models do not take into account such behavioral changes, realized health benefits may differ from anticipated benefits.

First, we explore the effect of the ECA adoption on campsite reservations using data from national park sites. Our outcomes of interest include the inverse hyperbolic sine of the number of campsite visits, days, and visitors.[19] We estimate,

$$Y_{imy} = \beta CMAQ_i \times postECA_{my} + \delta X_{imy} + \tau_{ry} + \alpha_{is} + \epsilon_{imy} \tag{5}$$

where i indexes a county in year-month ym. We include region-by-year, τ_{ry}, and county-by-season, α_{is} fixed effects. The other variables are defined analogously to equation (1). We also estimate results at the facility-year-month level that include region-by-year, facility-by-month, and year-by-month fixed effects.

Figure A11 shows the event-study style results from estimating equation (5). Prior to policy adoption in 2012, there is no evidence of differential pre-trends. After the policy began in 2012, there is a statistically significant increase in campsite reservations, and this increase is significant throughout the post-policy period. Columns (1), (3) and (5) of Table A11 show regression results from estimating equation (5) at the county level, while columns (2), (4), and (6) show facility level specifications including additional controls for facility-by-season and year-by-month fixed effects. A one-unit increase in CMAQ prediction is associated with a 14-16 percent increase in the number of visits, people, and days, after the ECA was implemented.

Next, we supplement these findings with data on time spent outdoors from the ATUS. Our outcome of interest is the inverse hyperbolic sine of total minutes spent outdoors.[20] We estimate,

$$Y_{jimy} = \beta CMAQ_i \times postECA_{my} + \delta X_{imy} + \pi Z_{ijmy} + \tau_{ry} + \alpha_{is} + \theta_{ym} + \epsilon_{ijmy} \tag{6}$$

where j indexes individuals in county i in year-month ym. The regression includes an additional set of individual-level controls, Z_{ijmy}, which include gender, race, ethnicity, education, age, pres-

[19] We use the inverse hyperbolic sine transformation rather than log transformation due to the presence of zeros. The inverse hyperbolic sine allows for the same interpretation as taking the natural log, but preserves zeros (Burbidge et al., 1988). Table A12 shows the results are robust to using a log transformation instead of the inverse hyperbolic sine.

[20] In Table A13 we report results for the extensive margin and intensive margin using a log specification that excludes zeros. The extensive margin estimates are only marginally insignificant (p-value=0.11) and indicate a 9 percent increase in spending any time outdoors. We further explore the distributional effects across bins of time spent outdoors in Table A14 where we observe shifts in both the lower and upper tails of the distribution. In the lower tail, there is a shift from zero minutes to between 0-1 hour. In the upper tail, we observe another shift from 3-5 hours to over 5 hours, indicating that for some individuals the gain in time outdoors may be larger.

ence of children in the household, and indicators for the day of the week of the survey and whether it was a holiday. The regression includes region-by-year (τ_{ry}), county-by-season (α_{is}), and year-by-month (θ_{ym}) fixed effects. Other variables are defined analogously to equation (1) and the regression is weighted using survey weights.

Figure A12 shows the event-study style results from estimating equation (6). There is no evidence of differential pre-trends prior to policy adoption. After the ECA was adopted, there was a gradual increase in time spent outdoors. Column (7) in Table A11 suggests that a one-unit increase in CMAQ prediction leads to an 8 percent increase in minutes spent outdoors. Relative to the baseline, this represents an increase of 1.2 minutes. In Table A15, we provide additional placebo tests on time spent on activities that are unlikely to be impacted by changes in air quality, such as sleeping, housework, and buying groceries. Reassuringly, we find no statistically significant impacts on these outcomes.

Across both datasets and a variety of measures, results suggest that policy-induced changes in air quality led to increased time spent outdoors. These behavioral changes can impact the reduced form effect of the ECA policy on health through decreased exposure to pollution or increased exercise, for example. Such complex behavioral changes make it especially important to quantify the health benefits of pollution regulation through ex-post policy evaluation.

6 Discussion & Conclusion

Policymakers frequently rely on the predictions of scientific models to anticipate the air quality improvement from a policy; yet, researchers infrequently test for differences between ex-ante and ex-post estimates. In this setting, only about half of the intended fine particulate matter improvements were realized, and we document evidence consistent with behavioral responses among shippers, other polluters, and individuals that are likely to contribute to deviations from the policy's anticipated impact. Our approach may be replicated in other settings with scientific research employing atmospheric transport model scenarios to improve estimation and policy evaluation.

Our results provide the first ex-post evaluation of US maritime emissions regulation. The US ECA led to meaningful improvements in fine particulate matter, infant health, and mortality as a result of maritime emissions controls. Combining CMAQ measurements with our estimated effect — that one unit of predicted fine particulate matter reduction from the ECA led to a 2 percent decline in low birth weight infants — and scaling by population, we calculate that the US ECA led to approximately 1,536 fewer low birth weight infants per year in areas near ship traffic. Similarly, we calculate that the policy resulted in a reduction of approximately 228 deaths per year under age one.[21] Using the EPA's value of a statistical life, this translates into $2.16 billion per year. The total benefits from improved health increase by an additional $139 million per year when we incorporate the effects of low birth weight on earnings (using estimates from Bharadwaj et al. (2018)) and the census bureau's work-life earnings.[22] The benefits to improved infant health alone are almost as large as the estimated cost of the policy, $3.2 billion in 2020.[23] Incorporating additional health benefits from cleaner air, such as fewer emergency room visits and hospitalizations, would likely lead to even higher total benefits. Moreover, research has shown that individuals respond to pollution information by exhibiting costly avoidance behaviors (Neidell, 2009; Deschenes et al., 2017; Keiser et al., 2018; Zhang and Mu, 2018). Our results on time spent outdoors provide evidence that individuals reduced avoidance in response to reduced air pollution from the ECA policy, so the health benefits alone can be considered a lower bound of the total benefits.

These findings are especially important given the IMO's recent adoption of a new global maritime sulfur emission standard in 2020, reducing sulfur content from 3.5 percent to 0.5 percent globally. Low-sulfur fuel can cost 30-50 percent more than bunker fuel, and fuel accounts for up to 75 percent of an ocean carrier's operation costs. This new regulation was estimated to cost the shipping industry between $10 to $60 billion per year depending on fuel prices (Corbett et al., 2016). Yet, our results suggest the potential for large benefits to human health in coastal areas throughout the world that have not yet adopted an ECA regulation. As of 2020, only the North American ECA, Baltic Sea ECA, and North Sea ECA were in effect. The health benefits from countries adopting the IMO's new global standard are likely to be quite large given that many countries had not regulated maritime sulfur emissions near coastal areas as of the time of our study.

[21] Figures A3 and A4 show the distribution of these health improvements spatially.

[22] We note that our estimates do not capture fetal deaths. To the extent that reduced air pollution from the policy decreased fetal deaths as well, the benefits we measure here are understated.

[23] To the extent that ship operators changed their routes to avoid using the more expensive low-sulfur fuel, the cost estimate may be an overestimate.

References

Alexander, Diane and Hannes Schwandt, "The Impact of Car Pollution on Infant and Child Health: Evidence from emissions cheating," *Review of Economic Studies*, 2022.

Arceo, Eva, Rema Hanna, and Paulina Oliva, "Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City," *The Economic Journal*, 2016, *126* (591), 257–280.

Auffhammer, Maximilian and Ryan Kellogg, "Clearing the air? The effects of gasoline content regulation on air quality," *American Economic Review*, 2011, *101* (6), 2687–2722.

Banzhaf, Spencer, Lala Ma, and Christopher Timmins, "Environmental justice: The economics of race, place, and pollution," *Journal of Economic Perspectives*, 2019, *33* (1), 185–208.

Barreca, Alan and Jessamyn Schaller, "The impact of high ambient temperatures on delivery timing and gestational lengths," *Nature Climate Change*, 2020, *10* (1), 77–82.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *The Quarterly Journal of Economics*, 2007, *122* (2), 775–805.

Becker, Randy and Vernon Henderson, "Effects of Air Quality Regulations on Polluting Industries," *Journal of Political Economy*, 2000, *108* (2), 379–421.

Bester, C Alan, Timothy G Conley, Christian B Hansen, and Timothy J Vogelsang, "Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators," *Econometric Theory*, 2016, *32* (1), 154–186.

Bharadwaj, Prashant, Petter Lundborg, and Dan-Olof Rooth, "Birth weight in the long run," *Journal of Human Resources*, 2018, *53* (1), 189–231.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes, "From the cradle to the labor market? The effect of birth weight on adult outcomes," *The Quarterly Journal of Economics*, 2007, *122* (1), 409–439.

Burbidge, John B, Lonnie Magee, and A Leslie Robb, "Alternative transformations to handle extreme values of the dependent variable," *Journal of the American Statistical Association*, 1988, *83* (401), 123–127.

Capaldo, Kevin, James J Corbett, Prasad Kasibhatla, Paul Fischbeck, and Spyros N Pandis, "Effects of Ship Emissions on Sulphur Cycling and Radiative Climate Forcing Over the Ocean," *Nature*, 1999, *400* (6746), 743–746.

Chay, Kenneth and Michael Greenstone, "Air Quality, Infant Mortality, and the Clean Air Act of 1970," *NBER Working Paper*, 2003, (w10053).

Corbett, James J, James J Winebrake, Edward W Carr, Jukka-Pekka Jalkanen, Lasse Johansson, Marje Prank, Mikhail Sofiev, SG Winebrake, and A Karppinen, "Health Impacts Associated with Delay of MARPOL Global Sulphur Standards," *IMO MEPC 70/INF*, 2016, *34.*

_ , _ , Erin H Green, Prasad Kasibhatla, Veronika Eyring, and Axel Lauer, "Mortality from Ship Emissions: A Global Assessment," *Environmental Science & Technology*, 2007, *41* (24), 8512–8518.

Currie, Janet, "Inequality at birth: Some causes and consequences," *American Economic Review*, 2011, *101* (3), 1–22.

_ **and Matthew Neidell**, "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?," *The Quarterly Journal of Economics*, 2005, *120* (3), 1003–1030.

_ **and Reed Walker**, "Traffic Congestion and Infant Health: Evidence from E-ZPass," *American Economic Journal: Applied Economics*, 2011, *3* (1), 65–90.

_ **, Matthew Neidell, and Johannes F Schmieder**, "Air Pollution and Infant Health: Lessons from New Jersey," *Journal of Health Economics*, 2009, *28* (3), 688–703.

Deschenes, Olivier, Michael Greenstone, and Joseph S Shapiro, "Defensive investments and the demand for air quality: Evidence from the NOx budget program," *American Economic Review*, 2017, *107* (10), 2958–2989.

Duflo, Esther, "The Economist as Plumber," *American Economic Review*, 2017, *107* (5), 1–26.

Eyring, Veronica, HW Köhler, J Van Aardenne, and A Lauer, "Emissions From International Shipping: 1. The Last 50 Years," *Journal of Geophysical Research: Atmospheres*, 2005, *110* (D17).

Fann, Neal, Karen Wesson, and Bryan Hubbell, "Characterizing the Confluence of Air Pollution Risks in the United States," *Air Quality, Atmosphere & Health*, 2016, *9* (3), 293–301.

Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth, "The effects of poor neonatal health on children's cognitive development," *American Economic Review*, 2014, *104* (12), 3921–55.

Fowlie, Meredith, Mar Reguant, and Stephen P Ryan, "Market-Based Emissions Regulation and Industry Dynamics," *Journal of Political Economy*, 2016, *124* (1), 249–302.

Gamper-Rabindran, Shanti and Christopher Timmins, "Hazardous waste cleanup, neighborhood gentrification, and environmental justice: Evidence from restricted access census block data," *American Economic Review*, 2011, *101* (3), 620–24.

Gillingham, Kenneth and Pei Huang, "Racial Disparities in the Health Effects from Air Pollution: Evidence from Ports," *NBER Working Paper*, 2021, (w29108).

Hellenic Shipping News, *2015 SOx Limits: Is The World Fleet Scrubbing Up?* 2014.

Hernandez-Cortes, Danae and Kyle C Meng, "Do Environmental Markets Cause Environmental Injustice? Evidence from California's Carbon Market," 2021.

Jacobson, Mark Z, *Atmospheric Pollution: History, Science, and Regulation*, Cambridge University Press, 2002.

Keiser, David, Gabriel Lade, and Ivan Rudik, "Air pollution and visitation at US national parks," *Science advances*, 2018, *4* (7), eaat1613.

Klotz, Richard and Julia Berazneva, "Local Standards, Behavioral Adjustments, and Welfare: Evaluating California's Ocean-Going Vessel Fuel Rule," *Journal of the Association of Environmental and Resource Economists (forthcoming)*, 2022.

Knittel, Christopher R, Douglas L Miller, and Nicholas J Sanders, "Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health," *Review of Economics and Statistics*, 2016, *98* (2), 350–366.

Lindgren, Samuel, "The coast is clear: Shipping emission standards, air quality and infant health," *Transportation Research Part D: Transport and Environment*, 2021, *100*, 103067.

Liu, Huan, Mingliang Fu, Xinxin Jin, Yi Shang, Drew Shindell, Greg Faluvegi, Cary Shindell, and Kebin He, "Health and Climate Impacts of Ocean-Going Vessels in East Asia," *Nature Climate Change*, 2016.

Luechinger, Simon, "Air Pollution and Infant Mortality: A Natural Experiment from Power Plant Desulfurization," *Journal of Health Economics*, 2014, *37*, 219–231.

Marcus, Michelle, "On the Road to Recovery: Gasoline Content Regulations and Child Health," *Journal of Health Economics*, 2017, *54*, 98–123.

Moore, Thomas J, Jessica V Redfern, Michael Carver, Sean Hastings, Jeffrey D Adams, and Gregory K Silber, "Exploring ship traffic variability off California," *Ocean & Coastal Management*, 2018, *163*, 515–527.

Moretti, Enrico and Matthew Neidell, "Pollution, Health, and Avoidance Behavior: Evidence from the Ports of Los Angeles," *Journal of Human Resources*, 2011, *46* (1), 154–175.

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

Oreopoulos, Philip, Mark Stabile, Randy Walld, and Leslie L Roos, "Short-, Medium-, and Long-Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins," *Journal of Human Resources*, 2008, *43* (1), 88–138.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Sanders, Nicholas and Alan I Barreca, "Adaptation to Environmental Change: Agriculture and the Unexpected Incidence of the Acid Rain Program," *NBER Working Paper*, 2021, (w28591).

Schlenker, Wolfram, "Daily Weather Data for Contiguous United States (1950-2019) - version March 2020," 2020. http://www.columbia.edu/~ws2162/links.html (accessed March 3, 2022).

_ **and W Reed Walker**, "Airports, Air Pollution, and Contemporaneous Health," *The Review of Economic Studies*, 2015, *83* (2), 768–809.

Smith, TWP, JP Jalkanen, BA Anderson, JJ Corbett, J Faber, S Hanayama, E O'Keeffe, S Parker, L Johansson, L Aldous et al., "Third IMO Greenhouse Gas Study 2014," 2015.

Sofiev, Mikhail, James J Winebrake, Lasse Johansson, Edward W Carr, Marje Prank, Joana Soares, Julius Vira, Rostislav Kouznetsov, Jukka-Pekka Jalkanen, and James J Corbett, "Cleaner Fuels for Ships Provide Public Health Benefits with Climate Tradeoffs," *Nature Communications*, 2018, *9* (1), 406.

Tessum, Christopher W, David A Paolella, Sarah E Chambliss, Joshua S Apte, Jason D Hill, and Julian D Marshall, "PM2.5 Polluters Disproportionately and Systemically Affect People of Color in the United States," *Science Advances*, 2021, *7* (18), eabf4491.

UNCTAD, "Review of Maritime Transport," Technical Report ISBN: 978-92-1-113073-7, United Nations, Geneva 2022.

U.S. EPA, "Frequently Asked Questions about the Emission Control Area Application Process," Technical Report, Office of Transportation and Air Quality, U.S. Environmental Protection Agency 2009.

_ , "Integrated Science Assessment for Particulate Matter," *US Environmental Protection Agency Washington, DC*, 2009.

_ , "Proposal to Designate an Emission Control Area for Nitrogen Oxides, Sulfur Oxides and Particulate Matter: Technical Support Document," Technical Report, Assessment and Standards Division, Office of Transportation and Air Quality, U.S. Environmental Protection Agency 2009.

_ , "Regulatory Announcement: Designation of North American Emission Control Area to Reduce Emissions from Ships," Technical Report, U.S. Environmental Protection Agency, Office of Transportation and Air Quality 2010.

_ , "Integrated Science Assessment for Oxides of Nitrogen - Health Criteria," *US Environmental Protection Agency Washington, DC*, 2016.

_ , "Integrated Science Assessment for Sulfur Oxides: Health Criteria ," Technical Report, U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment 2017.

_ , "Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI)," Technical Report, Air Quality Assessment Division, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency 2018.

Viana, Mar, V Rizza, Aurelio Tobías, E Carr, J Corbett, M Sofiev, A Karanasiou, G Buonanno, and N Fann, "Estimated health impacts from maritime transport in the Mediterranean region and benefits from the use of cleaner fuels," *Environment International*, 2020, *138*, 105670.

Wang, Chengfeng, James J Corbett, and Jeremy Firestone, "Modeling Energy Use and Emissions from North American Shipping: Application of the Ship Traffic, Energy, and Environment Model," *Environmental Science & Technology*, 2007, *41* (9), 3226–3232.

Winebrake, James J, JJ Corbett, EH Green, A Lauer, and V Eyring, "Mitigating the Health Impacts of Pollution from Oceangoing Shipping: An Assessment of Low-Sulfur Fuel Mandates," *Environmental Science & Technology*, 2009, *43* (13), 4776–4782.

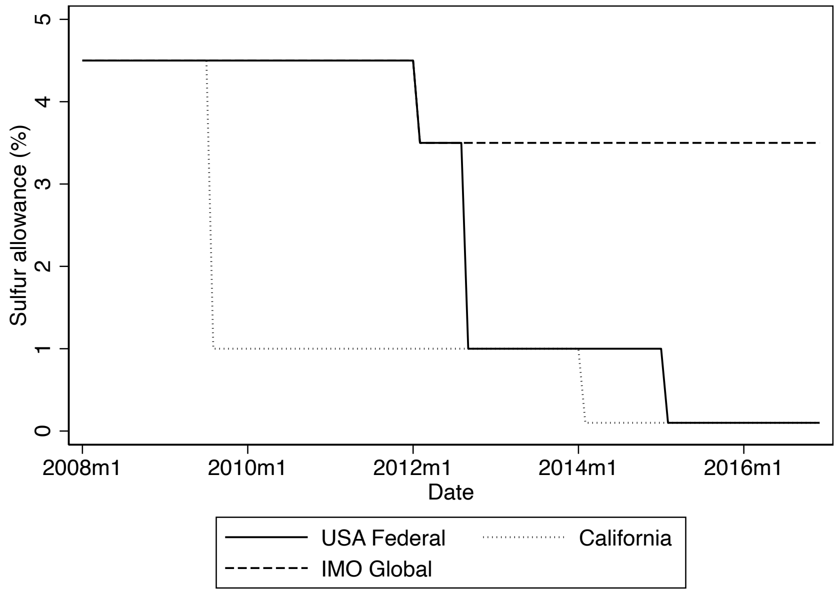
Zhang, Junjie and Quan Mu, "Air pollution and defensive expenditures: Evidence from particulate-filtering facemasks," *Journal of Environmental Economics and Management*, 2018, *92*, 517–536.

Zhu, Junming and Jiali Wang, "The effects of fuel content regulation at ports on regional pollution and shipping industry," *Journal of Environmental Economics and Management*, 2021, *106*, 102424.

Zou, Eric Yongchen, "Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality," *American Economic Review*, 2021, *111* (7), 2101–26.

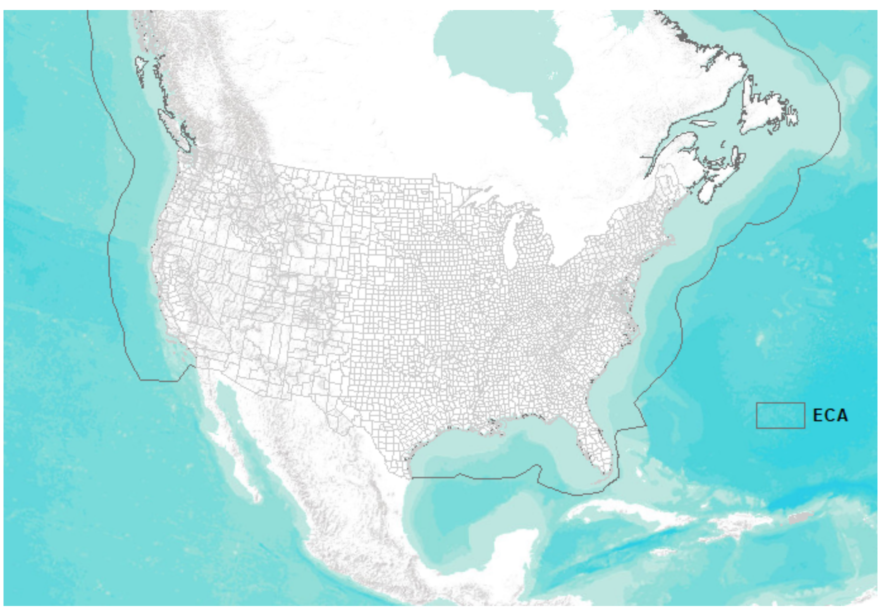
7 Figures

Figure 1: **Sulfur Allowance Limits Within and Outside Emission Control Areas**



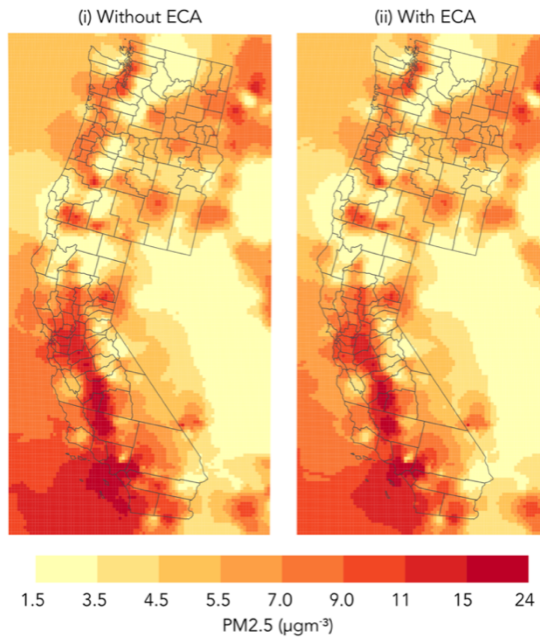
Note: California standard applies within 24 nautical miles of California's coast. USA federal standard applies within 200 nautical miles of coast. The International Maritime Organization (IMO) global standard applies elsewhere.

Figure 2: **North American Emission Control Area Boundary**

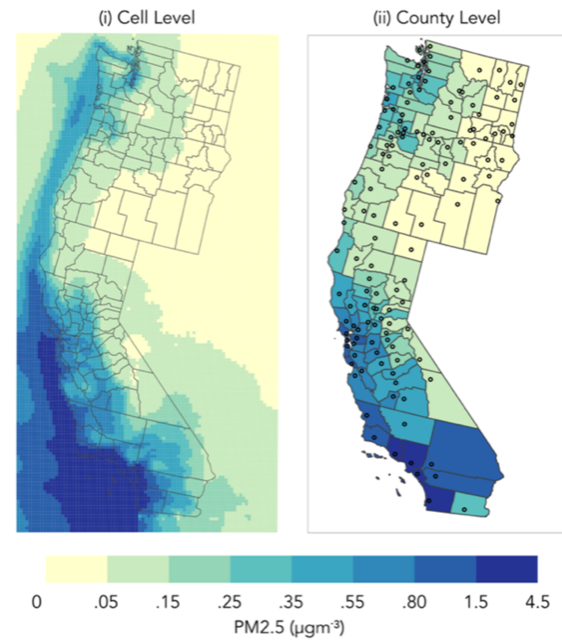


Note: Figure shows the regulated area for the North American Emission Control Area. Low sulfur fuel was required within the outlined boundary.

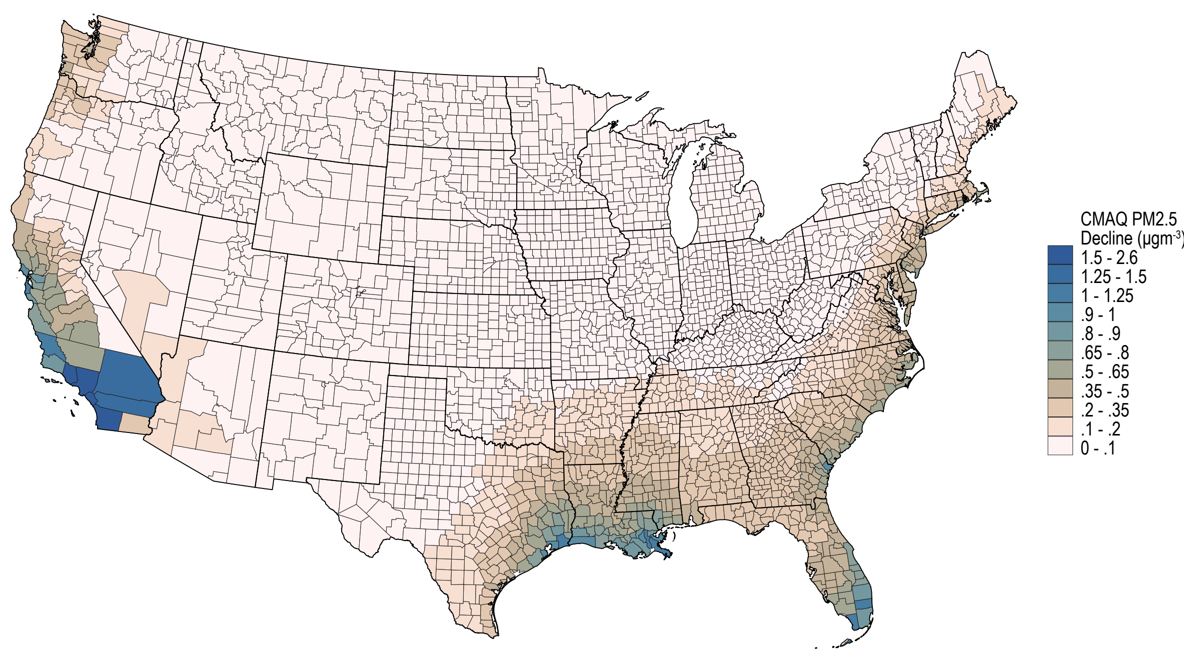
Figure 3: **CMAQ Predicted Decline in PM2.5 from ECA**



(a) CMAQ Output



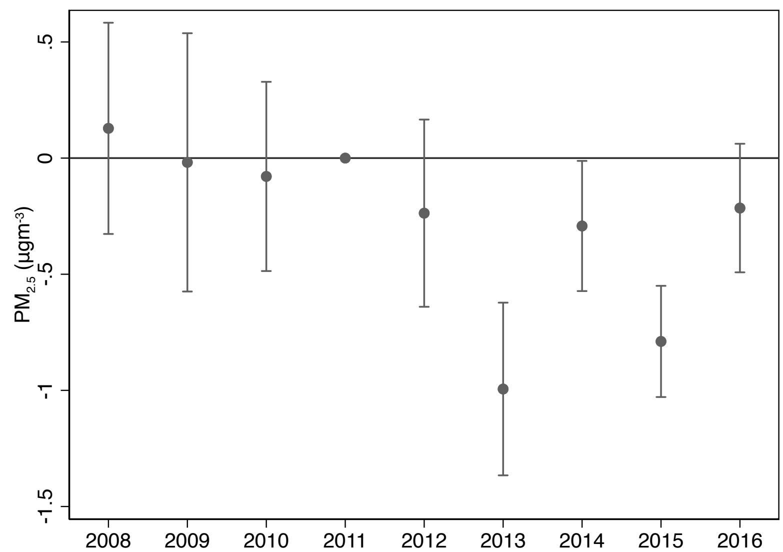
(b) CMAQ EPA Predicted Change



(c) CMAQ EPA Predicted Change for All Counties

Note: Figure shows the predicted decline in fine particulate matter from implementation of the North American Emission Control Area based on the CMAQ model. Panel (a) depicts the annual average ambient PM2.5 in 2020 under two CMAQ scenarios: (i) without the ECA policy and (ii) with the ECA policy. Panel (b) depicts the difference between the CMAQ scenarios of ambient PM2.5 in 2020 at (i) the cell level and (ii) the population-weighted centroid for a sub-sample of counties. Panel (c) depicts the CMAQ predicted change at the population-weighted centroid for all counties. Darker colors indicate a greater predicted decline in PM2.5. Data are from U.S. EPA (2009c).

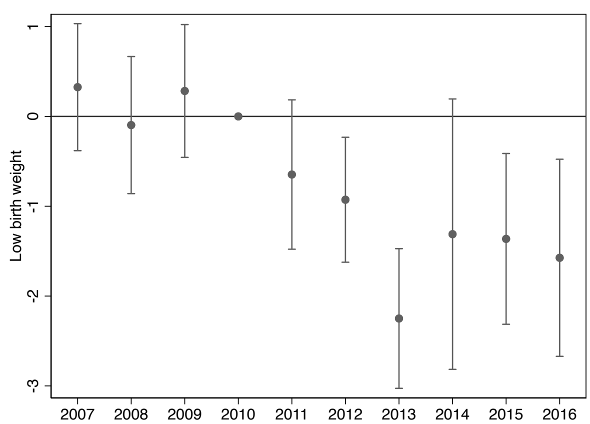
Figure 4: **Effects of ECA on Air Quality**



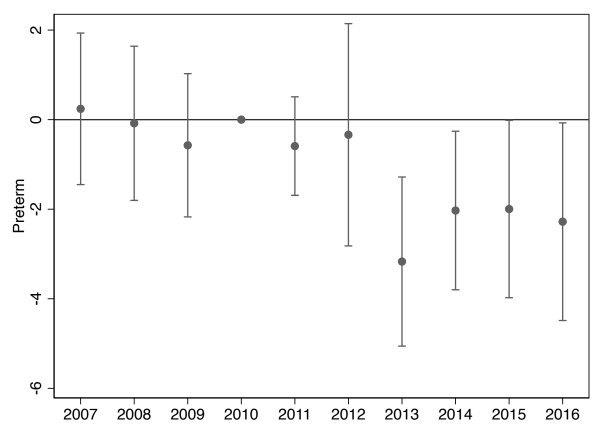
(a) PM2.5

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 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.

Figure 5: **Effects of ECA on Infant Health**



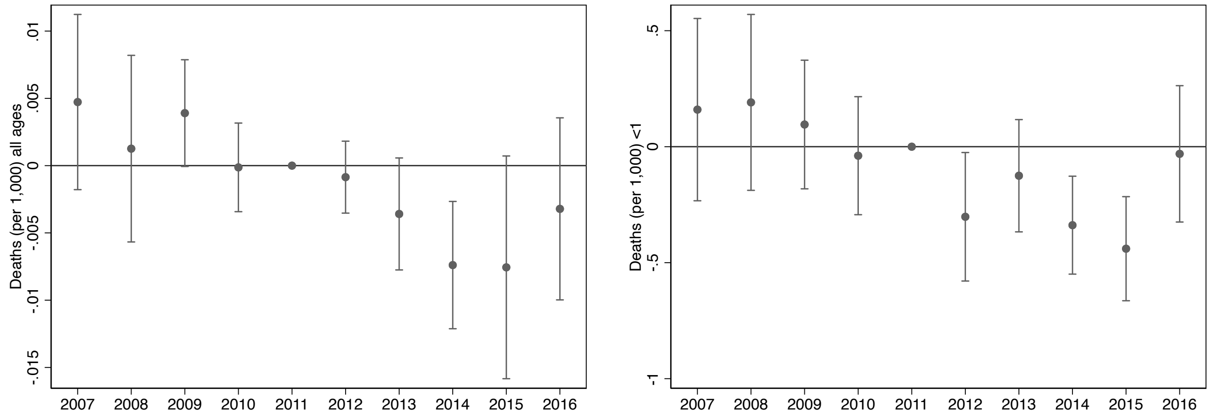
(a) Low Birth Weight



(b) Preterm Birth

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 depicted coefficients are the estimated effect of a one-unit increase in a county's CMAQ predicted reduction from the ECA in each conception year relative to 2010. The omitted period is conceptions in 2010 because conceptions in 2011 and 2012 were be partially treated. Conceptions in 2013 and after are fully treated during gestation. Robust standard errors are clustered at the county level. The confidence intervals are $\pm$ 1.96 standard errors.

Figure 6: **Effects of ECA on Mortality**

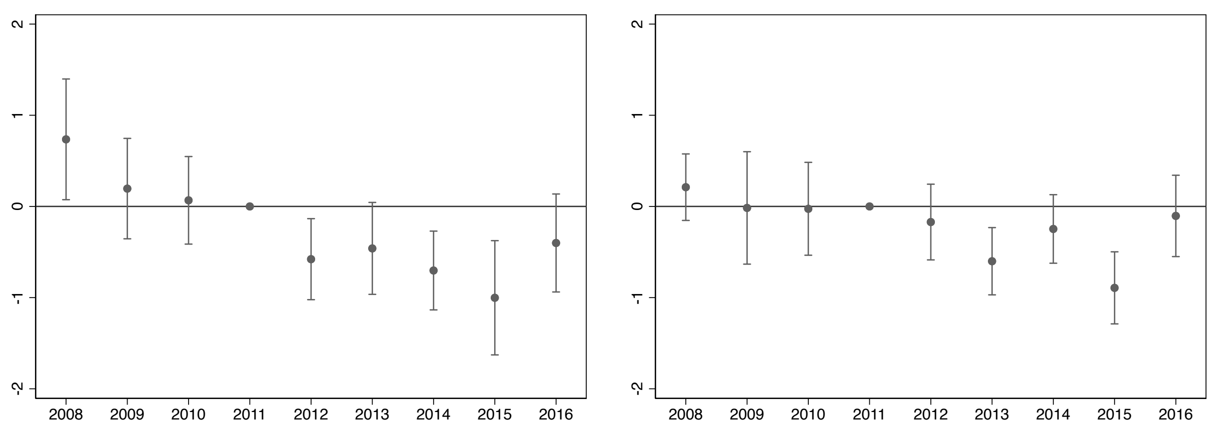


(a) All Ages

(b) Infant Mortality

Note: The unit of observation is a county-year-month. The observations are weighted by the total population in panel a and the number of births in panel b. 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 each year relative to 2011. Robust standard errors are clustered at the county level. The confidence intervals are $\pm$ 1.96 standard errors.

Figure 7: **Ship Behavioral Response: Full vs. Partial ECA**

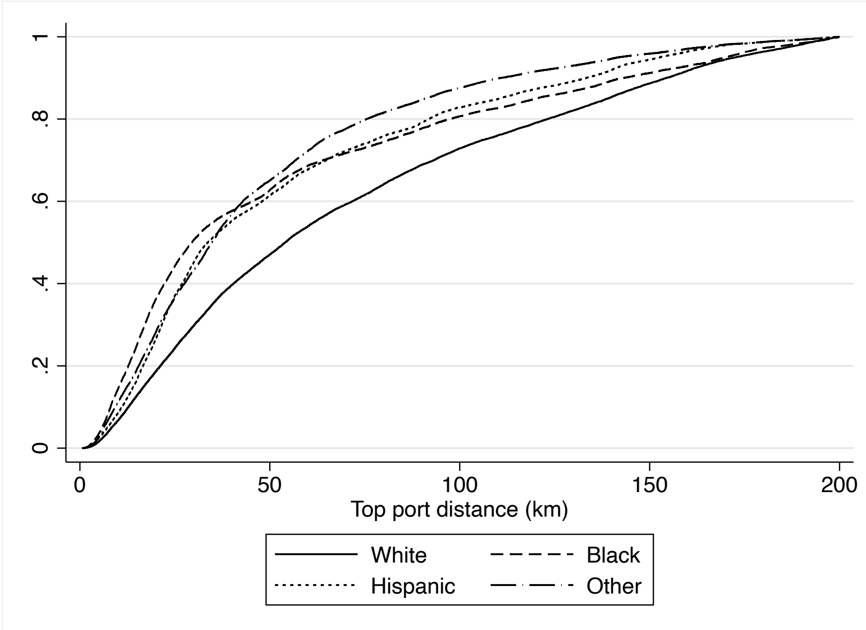


(a) Full ECA

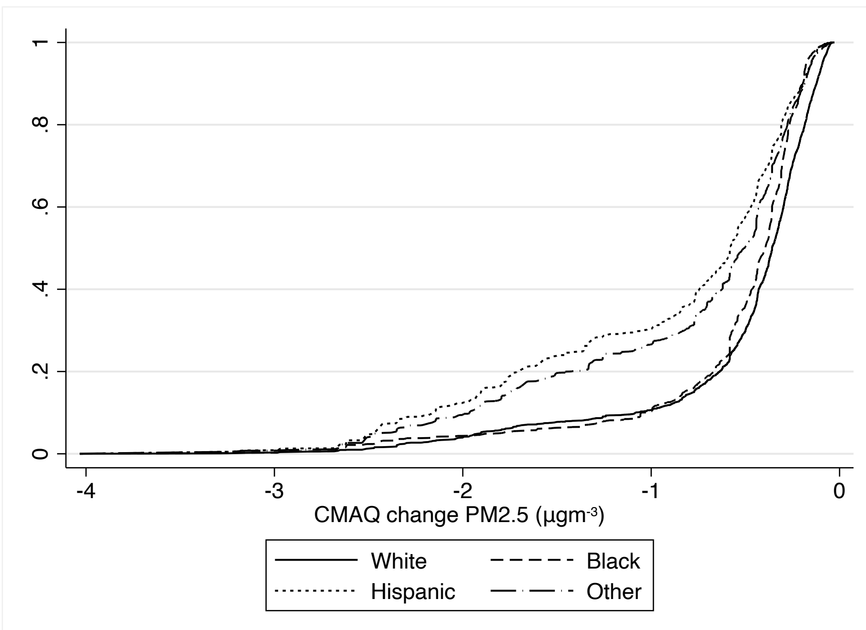
(b) Partial ECA

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 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.

Figure 8: **Disproportionate Exposure Among Populations Exposed to Maritime Pollution**



(a) Distance to Port



(b) CMAQ Exposure

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. Figure shows the cumulative distribution of individuals by race/ethnicity over distance to a port in panel (a) and intensity of ship emissions in panel (b). We calculate distance in kilometers from the population-weighted centroid of each tract to the nearest major port. The intensity of ship emissions is measured by the predicted change from requiring low sulfur maritime fuel from the CMAQ model at the centroid of each census tract.

8 Tables

Table 1: **Summary Statistics 2008-2016**

	All in 200k	In 200k and PM2.5 monitor
Outcomes		
PM2.5	9.25	9.25
Low birth weight (per 1,000)	61.61	60.53
Birth weight (g)	3,304.18	3,305.18
Pre-term (per 1,000)	95.55	93.82
Gestation(weeks)	38.77	38.78
Deaths (per 1,000) - Under 1	6.06	6.61
Deaths (per 1,000) - All	0.63	0.63
Mother characteristics		
Married	0.59	0.58
> HS Education	0.51	0.52
White	0.72	0.72
Hispanic	0.30	0.34
Over 35	0.18	0.19
Other controls		
Min temperature	9.87	9.88
Max temperature	21.41	21.40
Precipitation	2.80	2.64
Unemployment rate	7.72	7.85
Observations		
N conceptions/month	201.54	496.89
N counties	740.00	232.00

Note: The unit of observation is the county-year-month. The observations are weighted by the number of births conceived in county i in year-month ym. The sample in column 1 includes all counties with population-weighted centroids within 200km of heavy ship traffic. Column 2 drops counties without a PM2.5 monitor with at least one observation per year from 2008-2016. Means are reported for the main outcomes, demographic variables, and key control variables.

Table 2: **Effects of ECA on Air Quality and Demographic Characteristics**

	(1) PM2.5	(2) Maternal Index	(3) Log(conceptions)	(4) Unemp. rate
Post-ECA*CMAQ	-0.554	0.879	0.000	-0.108
	(0.104)***	(0.676)	(0.004)	(0.088)
R^2	0.60	0.95	0.99	0.93
N	24,901	25,052	25,052	25,052
N-counties	232	232	232	232
Mean	9.25	3305.18	2.33	7.85
%Change	-5.99	0.03	0.04	-1.38

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 in Columns 1, 2, and 4 and unweighted in Column 3. Reduced-form estimates are obtained from equation 1 with the modification of excluding the outcome variables as controls in columns 2 and 4. The first-stage impact on PM2.5 is reported in column 1. Column 2 reports the effect on an index of maternal characteristics, including education, marital status, race, ethnicity, age, smoking status, and diabetes. Column 3 repeats column 2 with the log number of conceptions as the outcome variable. Column 4 repeats column 2 with the unemployment rate as the outcome variable. The insignificant coefficients in Columns 2-4 indicate there is no evidence of changes in underlying economic and demographic characteristics that are correlated with the CMAQ policy variation. Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 3: **Effects of ECA and Air Quality on Health**

	(1) Low birth weight	(2) Preterm	(3) Deaths - all ages	(4) Infant Deaths
Panel A. Reduced Form				
Post-ECA*CMAQ	-1.326	-2.083	-0.006	-0.242
	(0.348)***	(0.782)***	(0.003)**	(0.089)***
R^2	0.57	0.63	0.92	0.63
N	25,052	25,052	25,056	25,052
N-counties	232	232	232	232
Mean	60.53	93.82	0.64	6.59
%Change	-2.19	-2.22	-0.92	-3.67
Panel B. 2SLS				
PM2.5	2.674	4.141	0.011	0.425
	(1.059)**	(2.126)*	(0.006)*	(0.199)**
R^2	0.45	0.53	0.91	0.61
N	24,901	24,901	24,905	24,901
F	18.33	18.33	30.40	27.37
N-counties	232	232	232	232
Mean	60.54	93.82	0.64	6.59
%Change	4.42	4.41	1.65	6.46
Panel C. OLS				
PM2.5	-0.004	0.016	0.002	0.009
	(0.036)	(0.055)	(0.000)***	(0.011)
R^2	0.57	0.64	0.92	0.63
N	24,901	24,901	24,905	24,901
N-counties	232	232	232	232
Mean	60.54	93.82	0.64	6.59
%Change Post-ECA	-0.01	0.02	0.29	0.14

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 conceptions (columns 1-2), population (column 3), and number of births (column 4). In Panel A, reduced-form estimates are obtained from equation 1. Panel B reports two-stage least squares estimates based on equation 2. Panel C reports the naive OLS estimates of pollution on health. The effects are reported for outcomes: low birth weight (<2,500g) per 1,000 (column 1), pre-term birth (<37 weeks) per 1,000 (column 2), deaths per 1,000 population (column 3), and infant deaths per 1,000 births (column 4). Robust standard errors clustered at the county level are reported in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 4: **Comparison of Magnitude to the Literature**

Study	Outcome	Pollutant	%Δ from 10% pollutant increase
Currie and Walker 2011	Low birth weight	NO2, SO2	17.65
Alexander and Schwandt 2020	Low birth weight	PM2.5, PM10, O3	10.3
H-L and Marcus	Low birth weight	PM2.5	4.1
Chay and Greenstone 2003 A	Infant mortality	TSP	5
Chay and Greenstone 2003 B	Infant mortality	TSP	3.5
Currie and Neidell 2005	Infant mortality	CO	1.01
Luechinger 2014	Infant mortality	SO2	0.89
Gutierrez 2015	Infant mortality	PM2.5, PM10	7.1
Knittel, Miller, Sanders 2016	Infant mortality	PM10	10.3
Alexander and Schwandt 2020	Infant mortality	PM2.5, PM10, O3	9.5
H-L and Marcus	Infant mortality	PM2.5	6.0

Note: Source of calculations from Alexander and Schwandt (2022).

9 Online Appendix

Contents

9.1 Data

9.1.1 Air Pollution Data

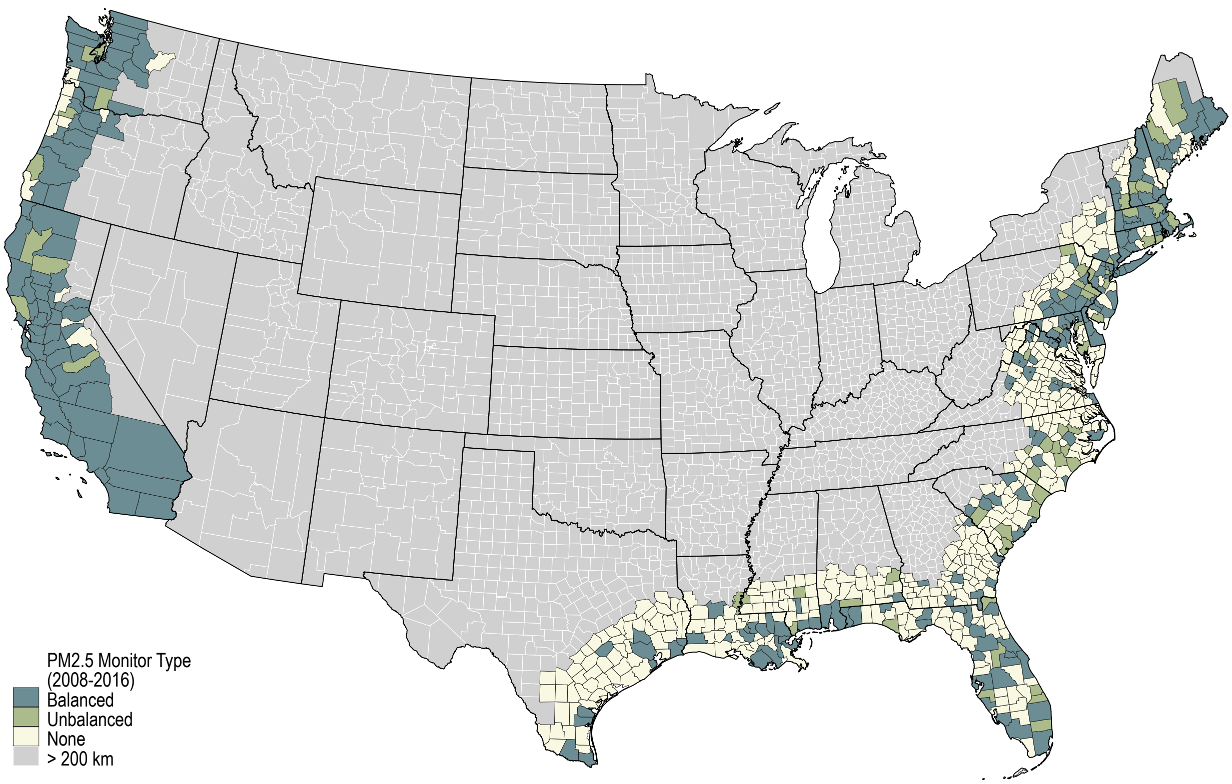
The source is the US EPA Air Quality System (AQS). The AQS records provide daily summaries from outdoor air quality monitors across the United States for a variety of pollutants.[24] Raw observations are at the level of the pollutant-monitor-day. We construct PM2.5-monitor-day observations from 3 PM2.5 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.

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,

[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 et al., 2016).

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.

9.1.2 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. 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.

9.1.3 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.

9.1.4 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.

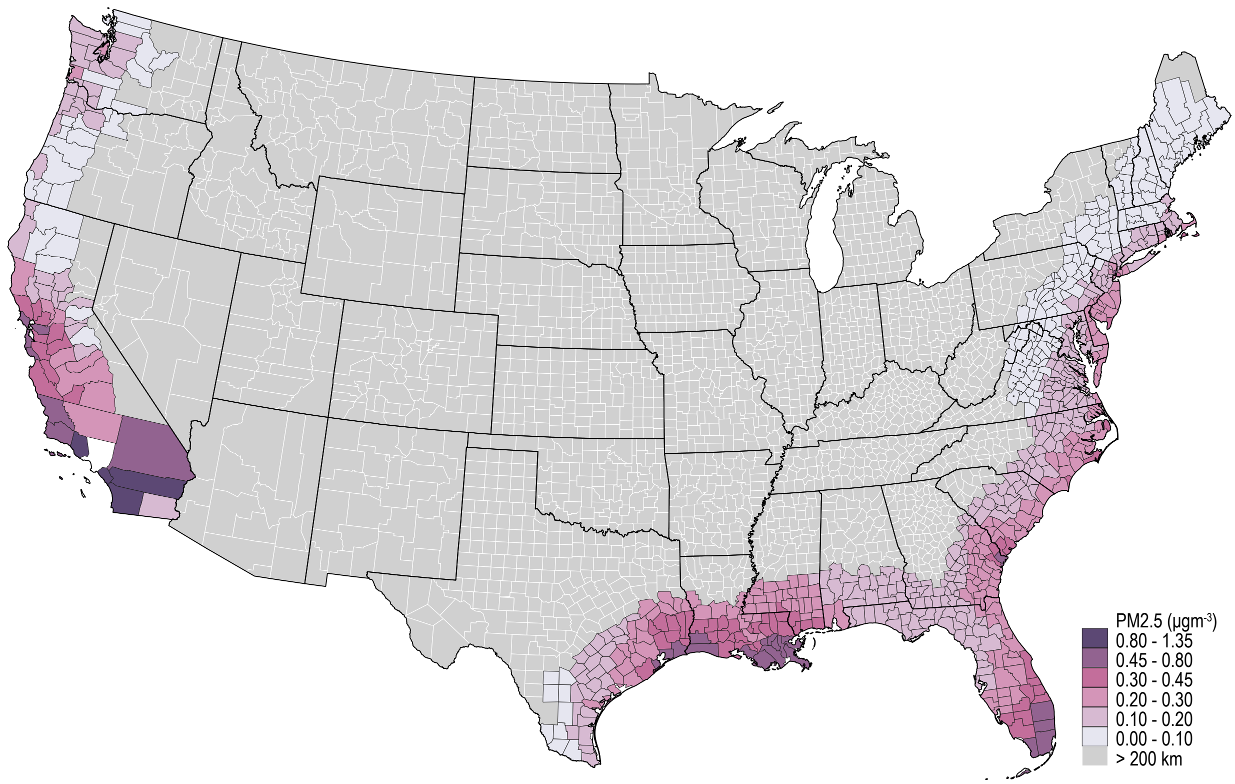
We supplement this with data from the American Time Use Survey (ATUS) from 2008 to 2016. Conducted by the US Census Bureau and the Bureau of Labor Statistics, the ATUS asks respondents to provide a detailed time diary of all activities over a 24-hour period, including the location of each activity. We 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.

9.2 ECA Effect Additional Results and Robustness

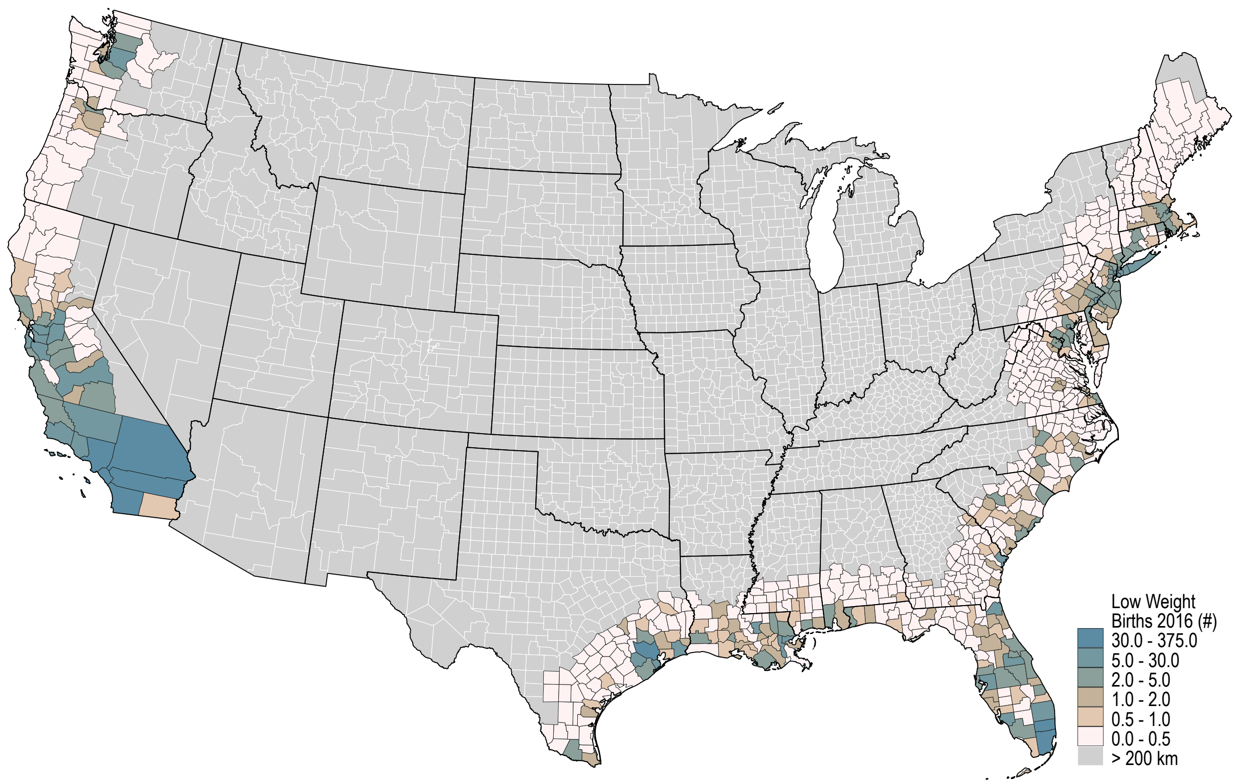
9.2.1 Spatial Distribution of Main Effects

Figure A2: Scaled Reduction in PM2.5



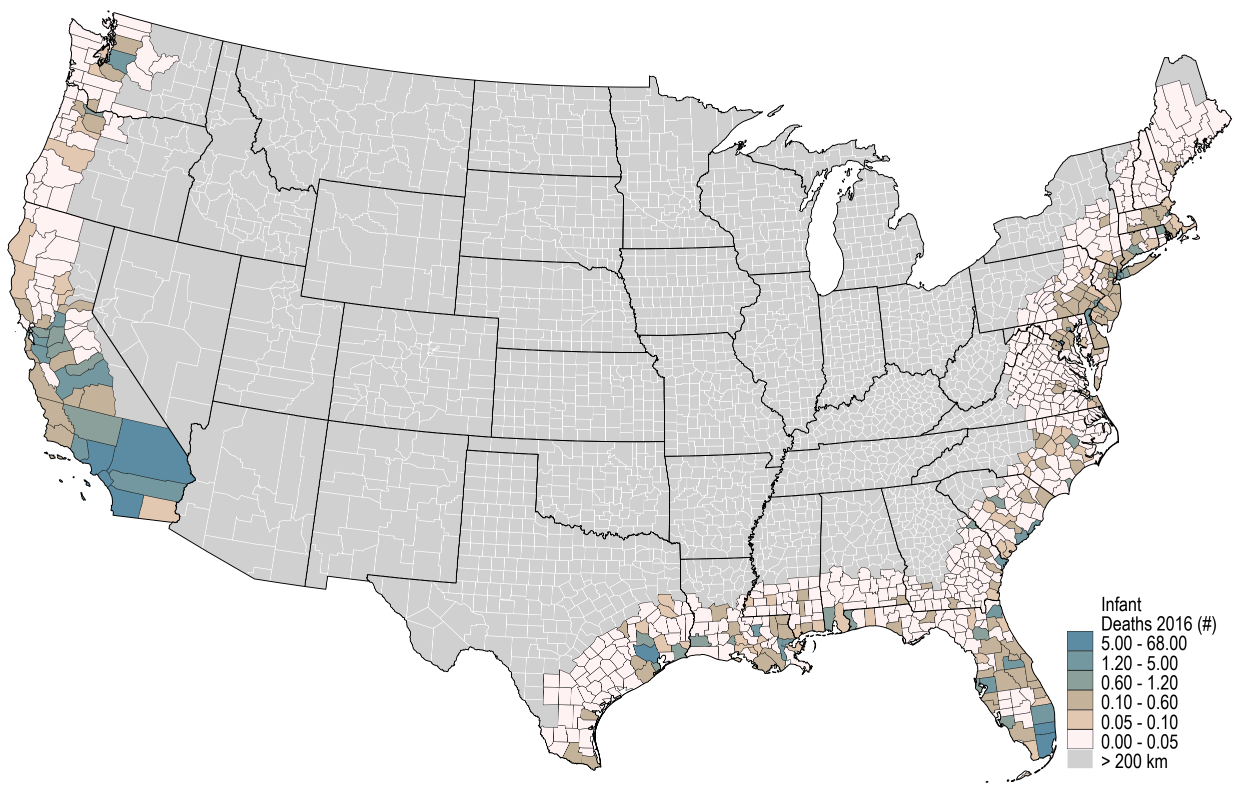
Note: Figure shows the estimated reduction in ambient PM2.5 from the ECA at the county level. The estimated reductions are the county level CMAQ predictions depicted in Figure 3 scaled by the estimated ECA effect coefficient in Table 2 column 1.

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.

9.2.2 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.

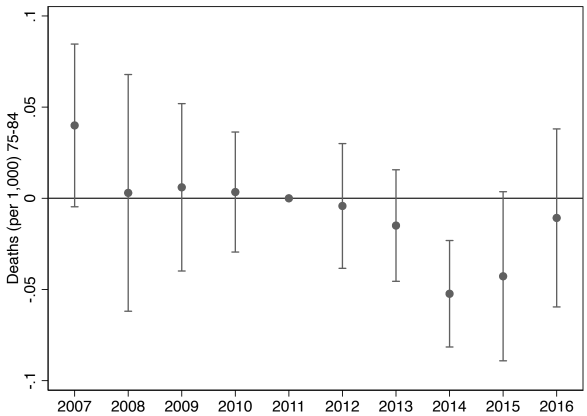
Similarly, We explore the distributional effect on birth weight further in the reduced form results in Panel A of Table A2, which shows the effect of the policy on bins of birth weight. Consistent with the stronger effects on infant health at the lower end of the distribution, we find large reductions 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**

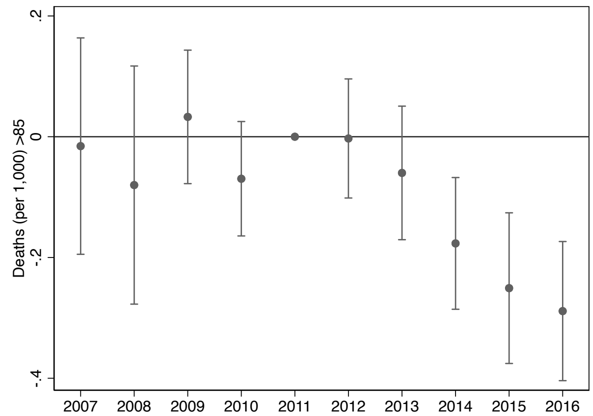
	(1) Birth weight	(2) Gestation	(3) Deaths: 75-84	(4) Deaths: >85
Panel A. Reduced Form				
Post-ECA*CMAQ	1.620	0.012	-0.031	-0.151
	(0.874)*	(0.006)**	(0.010)***	(0.043)***
R^2	0.82	0.71	0.77	0.65
N	25,052	25,052	25,056	25,056
N-counties	232	232	232	232
Mean	3305.18	38.78	3.72	10.90
%Change	0.05	0.03	-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.

9.2.3 Other Pollutants

We provide estimates of the impact of the policy on other pollutants in Table A3 below. 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: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

9.2.4 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.

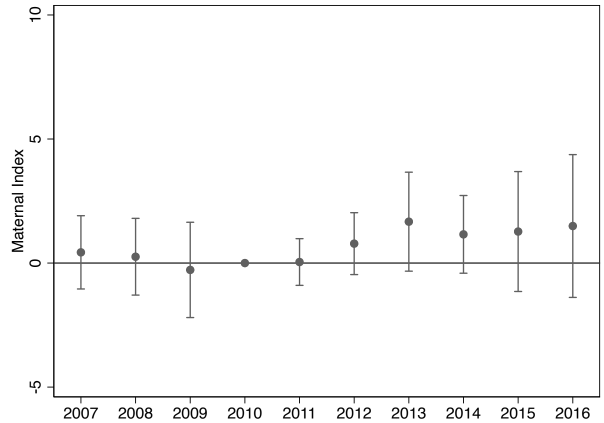
We failed to find evidence that demographic and economic shocks correlated with the treatment are driving the results in Table 2. To represent the many demographic characteristics relevant to infant health, we construct an index of maternal demographic characteristics, defined as the birth weight predicted 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), log emissions of PM2.5 (National Emissions Inventory), the number of other Toxic Release Inventory (TRI) pollution sources, 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)	(0.020)	(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.

9.2.5 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.

9.2.6 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. 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.

9.2.7 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: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

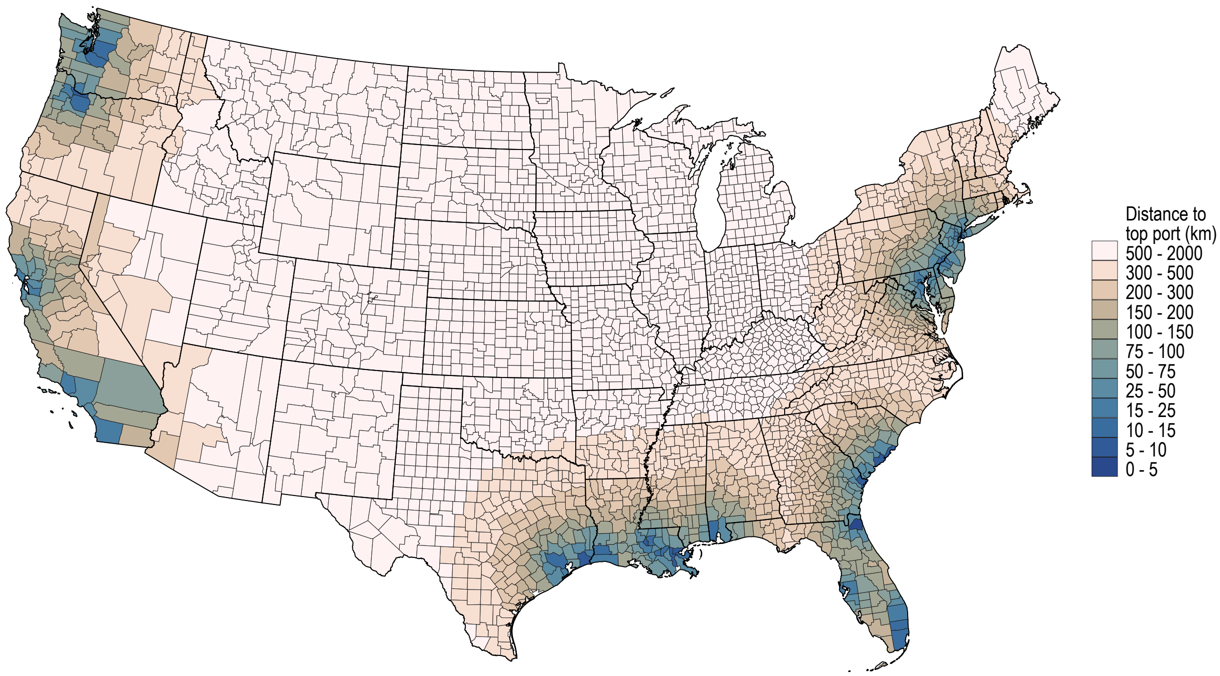
9.2.8 Comparing CMAQ with Distance Approach

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.

9.3 Incidence Results

9.3.1 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. 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.

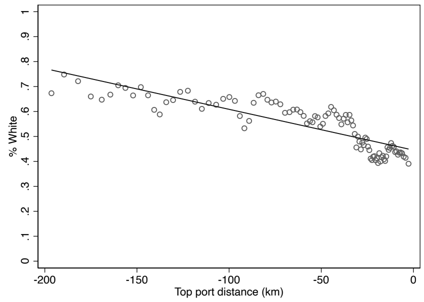
Figure 8 shows an alternative way to visualize these patterns. Here we present the cumulative distribution function of the proportion of individuals in each race/ethnicity group over distance to port (panel (a)) and intensity of ship emissions (panel (b)). A few interesting patterns stand out and are 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.

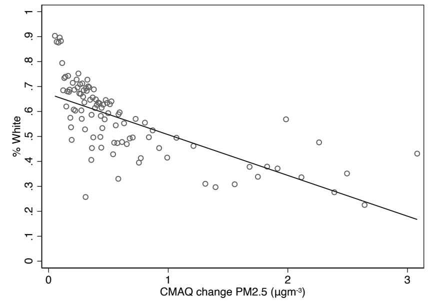
Given that the exposed population is different from most land-based pollution sources, the

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

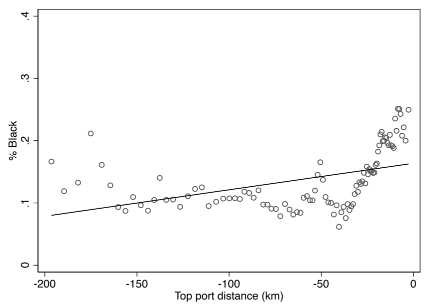
health effects of this policy are likely to be different than other reductions in air pollution. This is likely to be the case if, for example, pollution has a heterogeneous health impact across demographic groups, perhaps due to differences in underlying health conditions or access to care. In addition, 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.

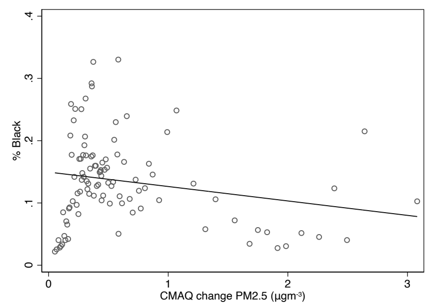
Figure A8: **Demographics of the Population Exposed to Maritime Pollution**



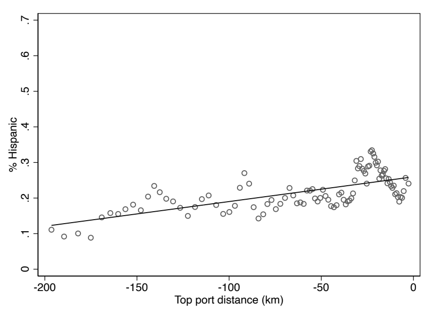


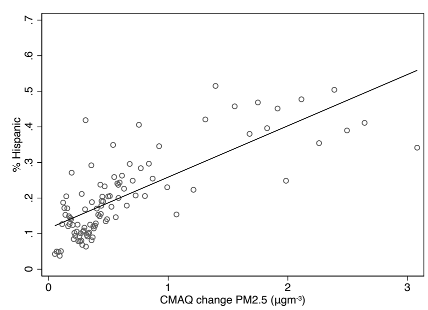
(a) Distance to Port & NH White (e) CMAQ Exposure & NH White



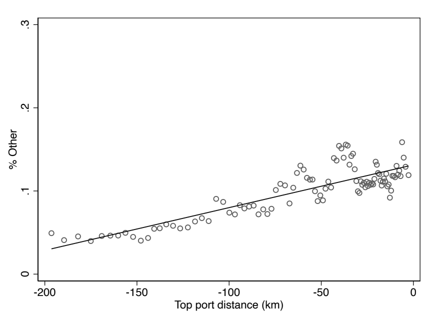


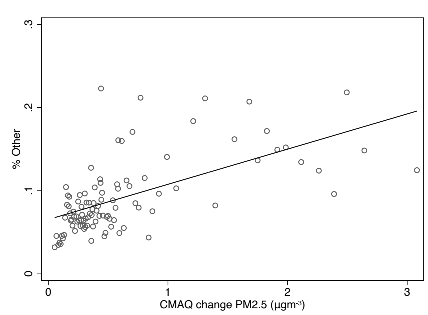
(b) Distance to Port & NH Black (f) CMAQ Exposure & NH Black





(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.

9.3.2 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: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

9.4 Behavioral Response Additional Results

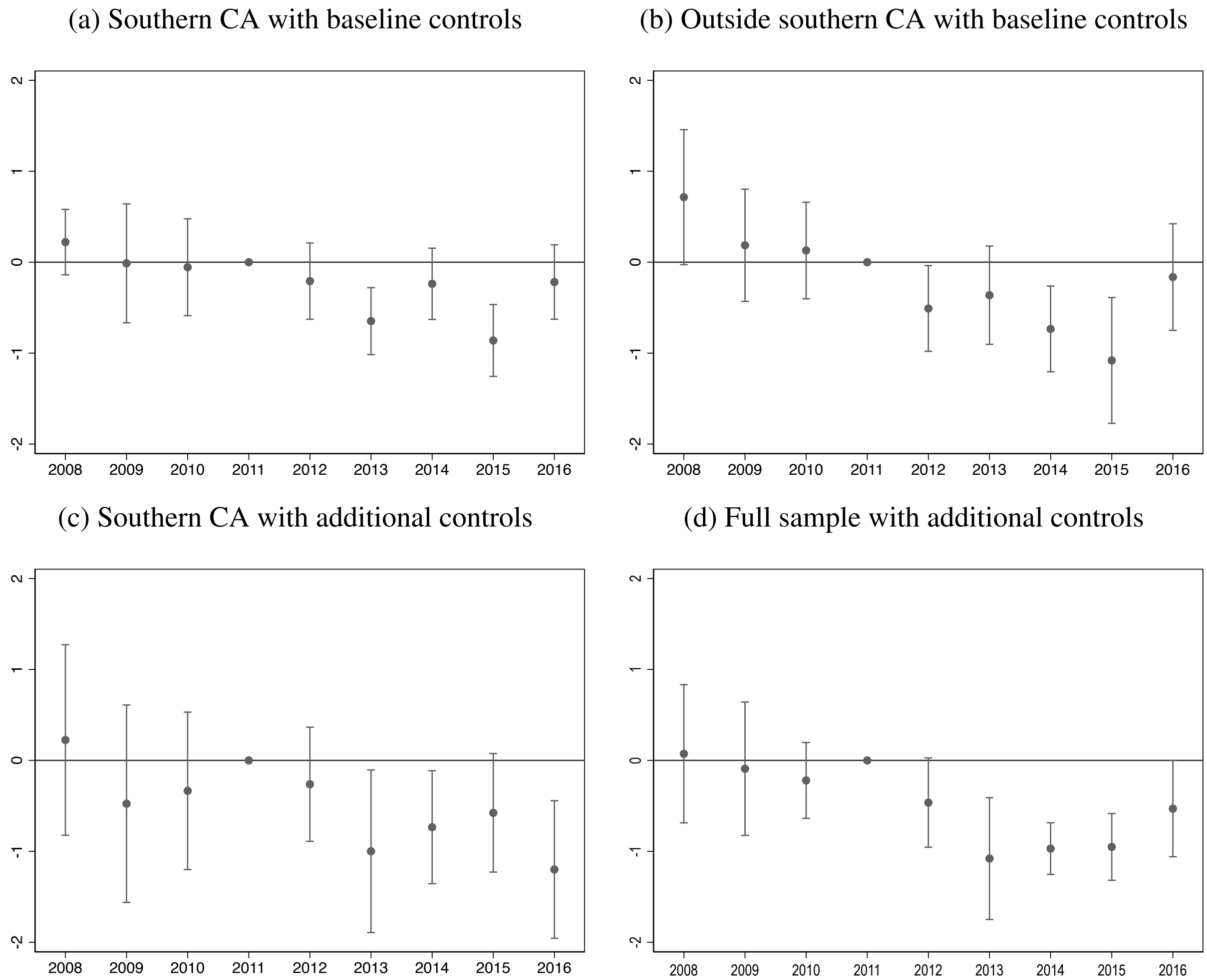
9.4.1 Ship Behavior

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.

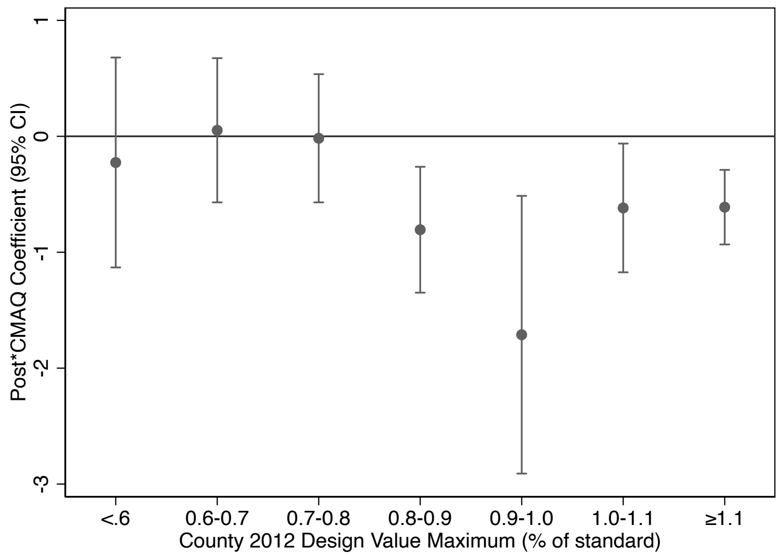
[31] We define southern California as the partial ECA counties in California. This includes nine counties: Imperial, Kern, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura.

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



9.4.2 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: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

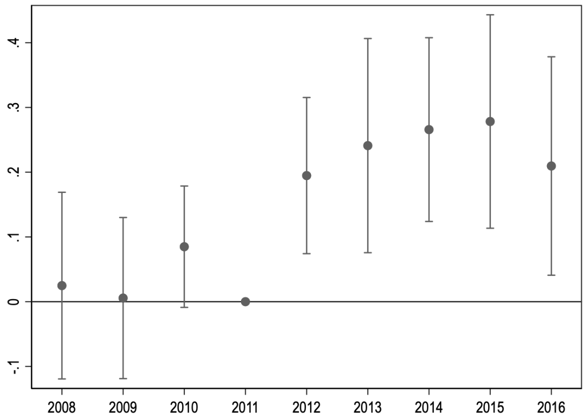
9.4.3 Individual Behavior

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

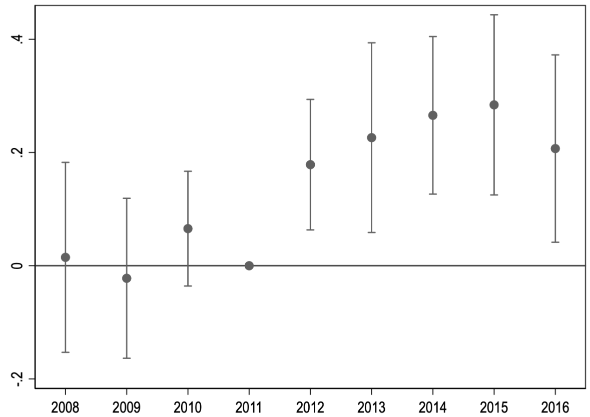
	Campsite Reservations (IHS)						Time Outdoors (IHS)
	Visits	Visits	People	People	Days	Days	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
post-ECA × CMAQ	0.164***	0.144***	0.149***	0.144***	0.166***	0.147***	0.0797*
	(0.0603)	(0.0325)	(0.0480)	(0.0357)	(0.0597)	(0.0307)	(0.0473)
Region-year FE	X	X	X	X	X	X	X
County-season 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**



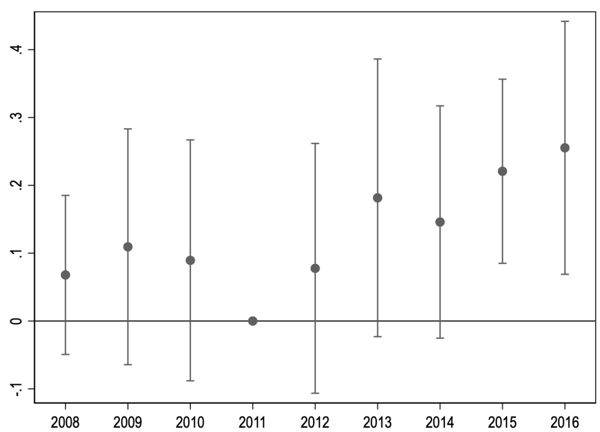
(a) Visits



(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 $\times$ 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 $\times$ 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).