Title
The Health Costs of Coal-Fired Power Plants in India

Authors
Barrows, Geoffrey
Garg, Teevrat
Jha, Akshaya

Publication Date
2019-12-10

Series Name: WPS
Paper No.: 101
Issue Date: 10 Dec 2019

The Health Costs of Coal-Fired Power Plants in India

Geoffrey Barrows, Teevrat Garg, and Akshaya Jha

CEGA
Center for Effective Global Action

Working Paper Series
Center for Effective Global Action
University of California

eScholarship
University of California

The Health Costs of Coal-Fired Power Plants in India

Geoffrey Barrows, Teevrat Garg, and Akshaya Jha*†

December 10, 2019

Abstract

This paper estimates the effect of coal-fired power plants on infant mortality in India. We find that a one GW increase in coal-fired capacity corresponds to a 14% increase in infant mortality rates in districts near versus far from the plant site. This effect is 2-3 times larger than estimates from the developed world. Our effects are larger for: (1) older plants, (2) plants located in areas with higher baseline levels of pollution, and (3) plants burning domestic rather than imported coal. The environmental benefits from policy aimed at the power sector are thus likely to be substantially higher if targeted at older plants located in more polluted areas tailored to burn domestic rather than imported coal.

JEL Codes: I15, Q51, Q56, Q48

Keywords: Coal, Electricity, India, Air Pollution, Infant Mortality, Infrastructure

*Geoffrey Barrows: Ecole Polytechnique, Email: geoffrey-masters.barrows@polytechnique.edu. Teevrat Garg: School of Global Policy and Strategy, UC San Diego, Email: teevrat@ucsd.edu. Akshaya Jha: H. John Heinz III College, Carnegie Mellon University, 4800 Forbes Avenue, Pittsburgh, PA 15213. Email: akshayaj@andrew.cmu.edu.

†We thank Prashant Bharadwaj, Karen Clay, Maureen Cropper, Lucas Davis, Tim Fitzgerald, Josh Graff Zivin, Lynne Kiesling, Dean Lueck, Leslie Martin, Craig McIntosh, Edson Severnini, and Frank Wolak, as well as seminar and conference participants at the 2017 AERE Summer Conference, the 2017 Annual Conference for the Society for Institutional and Organizational Economics (SIOE), Berkeley Energy Camp, Carnegie Mellon University, University of California - San Diego, University of Connecticut, University of Virginia, FGV-Rio, and the University of Illinois at Urbana-Champaign. We thank Carmen Sainz-Villalba, Kyle Navis and William Honaker for excellent research assistance. Garg acknowledges funding from the Center for Global Transformation and the Deep Decarbonization Initiative at UCSD. Any remaining errors are our own.

1 Introduction

Economic development is inextricably linked to increases in energy demand (Wolfram, Shelef and Gertler, 2012; Gertler et al., 2016). Developing countries have increasingly met this energy demand by building new coal-fired power plants.[1] Burning coal emits substantial quantities of local pollutants that harm the health of nearby populations (Graff Zivin and Neidell, 2013). It is thus imperative to assess the health costs of the rapid expansion of coal-fired capacity across the developing world.

Existing studies of the health costs of coal-fired capacity, and polluting industry more broadly, have focused primarily on the developed world (Luechinger, 2014; Clay, Lewis and Severnini, 2015; Currie et al., 2015; Beach and Hanlon, 2016; Cesur, Tekin and Ulker, 2017; Lavaine and Neidell, 2017; Johnson, LaRiviere and Wolff, 2017; Yang and Chou, 2017; Gibson, 2018). Estimates of the health costs of polluting industry from the developed world are not directly applicable to developing contexts for several reasons, including differences in baseline pollution levels, fuel burned, and production technologies (Hsiang, Oliva and Walker, 2019). However, in the absence of estimates specific to their country, policymakers in developing countries are often forced to calculate the environmental costs of polluting industry using crude extrapolations based on estimates from other more developed countries.

This paper estimates the health costs of coal-fired power plants in India. We focus on India for several reasons. First, India is home to over 1.2 billion people, making it the second most-populated country in the world. Moreover, India is rapidly building new coal-fired capacity to meet increasing electricity demand. Specifically, 75% of the grid-based electricity produced in India came from coal-fired sources, and this percentage could reach as high as 90% by 2030 (Shearer, Fofrich and Davis, 2017).[2] Third, India is one of the few developing countries with geographically disaggregated data on mortality. This allows us to more completely characterize heterogeneity in how coal-fired plants impact health. Fourth, migration rates in India are unusually low during our period of study (Munshi and Rosenzweig, 2009). Our estimates are thus unlikely to be biased by

[1]Between 1980 and 2018, the number of coal plants (total installed capacity) in the developing world increased from 180 (70 GW) to 1,563 (1,367 GW).

[2]India's planning commission released a report in 2017 stating that "...the reality of India's energy sector is that around three-quarters of our power comes from coal-powered plants and this scenario will not change significantly over the coming decades".

endogenous sorting in response to increases in coal-fired capacity. Finally, 14 of the 20 most polluted cities in the world are located in India. The health impacts of coal-fired power plants in settings with high baseline levels of pollution such as India potentially differ substantially from contexts with low baseline pollution levels (Arceo, Hanna and Oliva, 2016; Hsiang, Oliva and Walker, 2019).

We focus on the effect of coal-fired power plants on infant mortality rates because the largest costs from air pollution exposure are attributable to increases in mortality risk (EPA, 1999; Muller, Mendelsohn and Nordhaus, 2011) and infants are especially vulnerable to the adverse health impacts of air pollution exposure. Using data on the universe of coal-fired power plants in India from 1996-2014, we find that a 1GW increase in coal-fired capacity corresponds to a 14% increase in average district-level infant mortality rates. This effect is two to three times larger than comparable estimates from the developed world, and is comparable in magnitude to the infant mortality rates associated with deaths due to measles and malaria in India.

There is substantial heterogeneity how coal-fired capacity impacts IMR. Specifically, our results indicate that the estimated effect of coal-fired capacity on IMR is largest for older plants located in areas with high baseline pollution levels burning domestic coal rather than imported coal.[3] This heterogeneity has two direct policy implications. First, India should place more stringent environmental regulations on older plants located in areas with higher baseline pollution levels. This stands in contrast to current policies that apply more stringent standards to newer coal-fired plants - for example, the Clean Air Act of 1970 in the United States, which "grandfathered" older coal-fired plants in order to obtain political buy-in from the U.S. power sector (Hercher, 1980; Bushnell and Wolfram, 2012). Second, India is currently debating "protectionist" restrictions on coal imports (Varadhan, 2019). Our findings suggest that these import restrictions may come with substantial environmental costs if coal-fired plants burn domestic coal instead.

Our empirical specification relies on panel data variation in annual district-level changes in coal-fired capacity. Our specifications include controls for temperature and precipitation as well as district fixed effects and state-by-year fixed effects. We consider a battery of robustness checks. First, results from the event study framework formulated by Sandler and Sandler (2014) demonstrate that our estimated effect of coal-fired

[3]Coal sourced in India typically has higher ash content than coal imported from Australia, Indonesia or the United States.

capacity on IMR is not driven by pre-existing differences in trends for districts with and without changes in coal-fired capacity.[4] Second, we find no statistical impact of increases in non-coal-based (e.g.: hydro, nuclear, and natural gas) electricity production capacity on IMR. Third, we show that satellite-based measures of local air pollution increase in response to increases in coal-fired capacity but not increases in non-coal-fired capacity. Fourth, our estimated effects of coal-fired capacity on infant mortality rates and local air pollution are larger in magnitude downwind relative to upwind from the plant site (Herrnstadt and Muehlegger, 2015; Deryugina et al., 2016; Bondy, Roth and Sager, 2018).

We also test whether increases in coal-fired capacity impact economic outcomes in the district where the plant is located relative to other district in the same state (Matheis, 2016). We fail to find evidence of local economic benefits across a host of different outcome measures.[5] Specifically, our results indicate that there's no statistical difference in GDP or output and wages in the manufacturing and agricultural sectors in the district with the capacity increase relative to other districts in the same state. This suggests that, even in low-income countries with incomplete transmission grids, coal-fired power plants are a NIMBY. A direct implication of this finding is that plant siting decisions should be made primarily on the environment costs of the plant (which are local) rather than its economic benefits (which are not local).

Our paper makes two contributions to existing literature. First, as discussed above, previous research on the health costs of polluting industry have focused primarily on the developed world. To best of our knowledge, we provide the first econometric estimates of the effect of coal-fired power plants on infant mortality in a developing country. Closest to our paper, Gupta and Spears (2017) estimate the effect of coal-fired plants on respiratory health (coughing) in India. [6] In doing so, we can help inform prospective studies on the health benefits of proposed environmental policy in India (Cropper et al., 2012, 2019). In the absence of India-specific health estimates, Indian policymakers are often forced to assess the health costs of policy pertaining to coal-fired plants based on engineering estimates of how burning coal translates to pollution and pollution translates to health

[4]The method proposed by Sandler and Sandler (2014) accounts for differential timings of treatment as well as account for multiple instances of treatment in the same district.

[5]We fully acknowledge that our measures do not capture every possible benefit from coal-fired electricity production. However, these uncaptured benefits would have to be inordinately large in order to measure up against the health costs of coal-fired capacity.

[6]There is also a broader literature examining the effects of environmental regulation on health in developing countries (Greenstone and Hanna, 2014; Tanaka, 2015; Greenstone and Jack, 2015).

outcomes from other countries and industrial contexts.

Our second contribution is to document that the health costs of Indian coal-fired power plants are especially large for older plants located in areas with higher baseline pollution levels burning domestic rather than imported coal. This heterogeneity has important policy implications. For example, our results suggest that "grandfathering" older plants when implementing new environmental regulations comes with substantial environmental costs. Placing import restrictions on coal also causes significant environmental harm to the extent that coal-fired plants burn domestic coal instead. Finally, our results also indicate that coal-fired power plants do not come with significant local benefits relative to other districts in the same state. This suggests that policymakers should set plant-differentiated environmental policy or site new plants based primarily on the health costs of coal-fired plants rather than their (local) benefits.

The paper proceeds as follows. Section 2 discusses the institutional context surrounding the rapid expansion of coal-fired electricity generating capacity in India as well as our main data sources. We discuss the research design in Section 3. We show results pertaining to how power plant capacity impacts local air pollution and infant mortality rates in Section 4. Section 5 concludes by discussing the policy implications of our findings.

2 Institutional Background and Data

We consider the sample period 1996-2014. During this period, the annual total quantity of electricity consumed in India grew by 180%. India rapidly built coal-fired capacity to meet this rising electricity demand. Indeed, from 1996 to 2014, the number of coal-fired power plants more than doubled from 77 to 158 and total installed capacity roughly tripled, from 52GW to 156GW.[7] In 2014, over 75% of the grid-based electricity produced in India came from coal-fired power plants.

Appendix Figure A.1 presents the location of coal-fired power plants in India along with circles with 50km radius around each plant. The location of these plants has been subject to regulation as far back as the Third Five Year Plan (1961-66). Currently, new coal plants are required to be built further than 25km from the outer periphery of a city

[7] Appendix Figure A.2 plots the annual total number of plants and the annual total capacity by fuel type from 1939-2016.

and 500m away from the flood plain of any river system. The current guidelines thus reflect some portion of the environmental costs of siting coal-fired power plants in cities as well as the relatively low cost of transmitting electricity long distances. However, these guidelines are applicable only to newly-built plants. Many of the older coal-fired power plants still in operation today were built in or near cities in order to easily serve electricity demand in these cities.

Coal-fired power plants burn coal in order to heat water into the steam that drives the turbines used to produce electricity. Plants are thus typically sited near sources of water as well as either coal mines or coal transportation infrastructure. Finally, state and district boundaries also play a key role in plant siting decisions. This institutional context lends credence to our statistical analysis demonstrating that plant siting and capacity expansions are not tied to pre-existing trends in health and economic outcomes.

The vast majority of plants in India burn domestic coal; a small number of coastal plants burn coal imported from either Australia or Indonesia. Relative to coal mined in Australia or Indonesia, Indian coal typically has high ash content (ranging from 35-50%), high moisture content (4-20%), low sulfur content (0.2-0.7%), and low calorific values (between 2500-5000 kcal/kg) (Mittal, Sharma and Singh, 2012). As a result, burning Indian coal results in relatively low levels of SO_2 emissions, but the high moisture and ash contents along with the low heat content makes Indian coal particularly environmental unfriendly in terms of carbon dioxide (CO_2), nitrogen dioxide (NO_2), and fine particulate ($PM_{2.5}$) emissions.

There were no limits set on NO_2 and SO_2 emissions prior to December 2015.[8] In contrast, total suspended particulate (TSP) concentrations were regulated beginning in 1984 (Cropper et al., 2012). In addition, most coal-fired plants in India have installed electrostatic precipitators (ESPs) designed to reduce $PM_{2.5}$ emissions. Consequently, our analysis of the pollution concentration levels around plant sites focuses on NO_2.

2.1 Data on Air Quality

Unreliable and sparsely distributed ground monitors have historically limited empirical research on air pollution in the developing world (Donaldson and Storeygard, 2016).

[8]India did set requirements on the minimum height of power plant smokestacks, which impacts the extent to which emissions are dispersed across space.

However, with the advent of retrospective analysis, atmospheric scientists have developed methods to convert satellite readings of aerosol optical depth (AOD) into gridded pollution concentration products. We utilize two such products.

First, Van Donkelaar et al. (2016) constructs annual NO_2 concentrations across the entire world at a 0.1° × 0.1° resolution for 1996 - 2015. Similarly, Van Donkelaar et al. (2016) constructs annual $PM_{2.5}$ data gridded at the 0.01° × 0.01° resolution for 1998-2015. Our second source of pollution data is the Modern-Era Retrospective analysis for Research and Applications (MERRA) database. MERRA lists monthly $PM_{2.5}$ and SO_2 for the sample period 1980-2016 gridded at the 0.5° × 0.625° resolution. Descriptive statistics for all of our measures of pollution are reported in Table A.2.[9]

2.2 Data on Infant Mortality

The cornerstone of our analysis is annual data from the Vital Statistics of India on infant mortality rates (IMR). This data-set contains annual district-level information on number of infant deaths and number of live births in each year in each district.[10] The Vital Statistics of India is the best available data on district-level IMR that spans all of India (Greenstone and Hanna, 2014; Burgess et al., 2014). That being said, many births and deaths go unreported in practice. As discussed in Greenstone and Hanna (2014) and Burgess et al. (2014), under-reporting is more prevalent for deaths than births, but the year-to-year variation in IMR looks similar in our data relative to other survey-based data-sets.

Our data separately reports annual district-level live births and infant deaths for urban versus rural areas within the district. For each category – total, urban, and rural – we compute IMR as number of infant deaths per 1,000 live births. Unfortunately, we do not know the geographic borders used to determine this urban versus rural classification, so we cannot determine whether a coal-fired power plant would be classified as being in an "urban" versus "rural" area.[11] That being said, we separately estimate the impact of

[9]Satellite-based measures of NO_2 are significantly lower than ground monitor readings on average because Van Donkelaar et al. (2016) measures the entire NO_2 "column" while air quality monitors record ground-level NO_2 concentrations (Bechle, Millet and Marshall, 2013). This distinction is not important for our goal of demonstrating that there's an economically and statistically significant effect of coal-fired capacity on NO_2 concentration levels at different distance bandwidths from plant sites.

[10]Infants are defined in the data as children ages 1 or less.

[11]Importantly, the classification is not urban versus rural districts but rather urban versus rural areas

annual district-level coal-fired capacity on total, urban and rural infant mortality rates. We report descriptive statistics in Appendix Table A.2.

3 Research Design

In this section, we describe the construction of measures of district-level electricity production capacity as well as the subsequent research design employed in estimating the effects of production capacity on local air pollution and infant mortality rates.

3.1 Measures of Electricity Production Capacity

The main results in this section relate variation in coal-fired electricity generating capacity to district-level air pollution and infant mortality rates. We compute two different measures of coal-fired capacity. First, we simply sum the installed capacity of all of the coal-fired plants in each district d in each year t:

$$Cap_{d,t} = \sum_{p=1}^{P} PlantCap_{p,t} \tag{1}$$

where p indexes plant.

However, previous research indicates that the pollution emissions from coal-fired power plants can travel hundreds of kilometers (Muller, 2014). As shown in Figure A.1, this implies that the airborne emissions from coal-fired power plants can easily cross district borders. These cross-district flows of pollution may bias estimates of the effect of coal-fired capacity on health outcomes towards zero, as even "control" districts without coal-fired power plants will be impacted by pollution from coal-fired plants. Consequently, we construct an annual district-level measure of electricity generating capacity that accounts for both spillovers across districts and wind direction from the plant.

We incorporate wind direction as part of this measure of capacity because of a growing literature that utilizes wind direction to identify how local air pollution affects economic and environmental outcomes (Anderson (2015); Herrnstadt and Mueh-

in the same district. As discussed in Appendix Section B.1, most coal-fired power plants tend to be located in more populated, urban areas in a district.

legger (2015); Deryugina et al. (2016)). We calculate area-based measures of annual district-level capacity separately for plants upwind versus downwind from the district. In particular, we compute the monthly average wind direction from each power plant in our data using the Modern-Era Retrospective analysis for Research and Applications (MERRA) data-set provided by NASA. We then construct a quarter-circle with radius 100km being intersected with the arc 45 degrees less than and 45 degrees greater than the wind direction provided by MERRA; this wind direction is flipped by 180 degrees when constructing "upwind" capacity. Next, we calculate the share of the district that is covered by each of these quarter circles, multiply by the capacity of the plant, and sum over plants:

$$CapWind_{d,t}^{radius=100,DW} = \sum_{p=1}^{P} PlantCap_{p,t} \times \frac{PlantQuarterCircle_{p,d}^{radius=100,DW}}{DistrictArea_d} \tag{2}$$

where $\frac{PlantQuarterCircle_{p,d}^{radius=100,DW}}{DistrictArea_d}$ is the share of district d that is covered by a quarter circle downwind of plant p in month t. Upwind capacity is computed similarly.

Though the previous discussion has centered on coal-fired capacity, we construct both the in-district and wind-and-area-based measures of capacity for other types of power plants as well. Specifically, we aggregate all power plants with fuel type other than coal (nuclear, gas, and hydro) into a single category, calculating annual district-level measures of "non-coal" capacity. Descriptive statistics are reported in Appendix Table A.1.

3.2 Empirical Specifications

We consider a panel regression model relating each of our measures of capacity on air pollution and infant mortality rates. We consider four different measures of capacity $\text{CapMeasure}_{d,t}$ for each district d in each year t: coal- versus non-coal fired capacity, measuring capacity based on plants in the district versus wind-direction-informed cones around plants. For each of these measures, we estimate the following specification for outcome $Y_{d,t}$.

$$Y_{d,t} = \alpha_d + \theta_{s,t} + \beta \text{CapMeasure}_{d,t} + X_{d,t}\gamma + \epsilon_{d,t} \tag{3}$$

We control for temperature and precipitation (i.e.: $X_{d,t}$). We obtain weather data from ERA Interim at the 1 degree by 1 degree resolution and aggregate up to the district-year level (Schlenker and Lobell, 2010; Auffhammer et al., 2013). This specification also controls for district fixed effects α_d and state-by-year fixed effects $\theta_{s,t}$.

We statistically test for differential pre-existing trends across districts with versus without capacity increases using the event study framework formulated in Sandler and Sandler (2014). The method described in Sandler and Sandler (2014) allows for multiple "events" per district, where an "event" in our context is any increase in the level of capacity in the district. It is important to account for the possibility of multiple events because more than one plant can be built in a district and some plants install new capacity multiple times over our 1996-2014 sample period.

Formally, we estimate:

$$Y_{d,t} = \alpha_d + \theta_{s,t} + \sum_{j=1}^{J_d} \sum_{m=-M}^{M} \beta_m 1(t - e_{d,j} = m)(\Delta Cap)_{d,t} + X_{d,t}\gamma + \epsilon_{d,t} \quad (4)$$

for each event j occurring in year $e_{d,j}$ (ex: an increase in coal-fired capacity in district d in year $e_{d,j}$). Each coefficient β_m captures the impact of an event on outcome $Y_{d,t}$ m years in the past or future. If the timing of plant openings and capacity additions is exogenous to unobserved determinants of the outcome variable (i.e.: no pre-existing differential trends), $\beta_m = 0$ should be zero all $m < 0$. For ease of presentation, we bin event years into the following categories: between 1 and 5 years before the event, zero to four years after the event, five to nine years after the event, and more than 10 years after the event. The excluded category is more than 5 years before the event, and we do not report the coefficient estimate for more than 10 years after the event because this bin contains relatively few observations (i.e.: it's an "endpoint restriction" bin). Standard errors are clustered by state for all of our specifications.

4 Local Health Costs of Coal-Fired Power Plants

This section presents the effect of coal-fired power plants on air pollution and infant mortality rates (IMR) in India. We discuss heterogeneity in the effect of coal-fired capacity on IMR in the third subsection. We conclude by discussing a host of sensitivity analyses, including tests for differential pre-trends as well as placebo tests based on non-coal power plants.

4.1 The Effect of Coal-Fired Capacity on Air Pollution

In Tables 1 and A.3, we estimate Equations (3) and (4) taking the log of district average pollution measures as the dependent variable. In presenting these regressions, we aim only to demonstrate that there exists a plausible channel through which coal-fired capacity can affect infant health. Of course, coal-fired plants may adversely impact health through other channels such as groundwater pollution as well. We estimate the impacts of coal-fired plants on pollution in Columns 1-4, and the impacts for non-coal plants in Columns 5-8.

[Table 1 about here.]

Focusing first on Column 1 of Table 1, a 1 GW increase in coal-fired capacity increases district-level average NO_2 by 7.6%. This effect is statistically significant at the 1% level.[12] In contrast, there is no statistical response in average NO_2 levels in response to increases in non-coal capacity (see Column 5 of Table 1). This is precisely what we should expect: producing electricity using gas-fired, nuclear or hydro sources does not emit significant amounts of NO_2. Finally, we fail to reject the null hypothesis of no differential pre-existing trends prior to capacity increases. To see this, note that the coefficient estimate corresponding to between 1 and 5 years before a capacity increase is small and not statistically different from zero in Column 1 of Panel B of Table 1.

Across multiple satellite products, we find that average concentration levels of fine particulates ($PM_{2.5}$) do not vary with increases in coal-fired capacity. At first glance, this may seem somewhat surprising, since a large literature connects $PM_{2.5}$ to health outcomes, and burning coal emits $PM_{2.5}$. There are two potential explanations for this null result. First, coal-fired plants are not a major source of $PM_{2.5}$ in India, perhaps because most coal-fired plants in India are equipped with technology designed to abate PM emissions. In fact, according to the national emissions inventory, coal-fired plants only account for 20% of PM emissions nationwide. By contrast, coal-fired plants account for 60% and 55% of NO_2 and SO_2 emissions respectively. Hence, there may be too much background noise to detect any coal-plant-related increases in $PM_{2.5}$. Second, $PM_{2.5}$ is

[12] An extant literature in epidemiology and economics documents the relationship between NO_x exposure (often through conversion to ozone) and human health (Wolfe and Patz, 2002; Lleras-Muney, 2010; Mölter et al., 2014; Deschenes, Greenstone and Shapiro, 2017).

harder than NO_2 to detect because $PM_{2.5}$ is a mixing pollutant (Grainger, Schreiber and Chang, 2018).

In the Appendix, we consider two additional sensitivity analyses pertaining to the impact of coal-fired capacity on NO_2 levels. First, Appendix Table A.3 demonstrates that the effect of coal-fired capacity on NO_2 is larger for districts downwind from plant sites relative to upwind from these sites on average. Second, we assess how far the NO_2 emissions from coal-fired plants travel using a spatial difference-in-differences framework in Appendix Table A.4. Comparing the estimates across columns, we detect a statistically significant effect of coal-fired capacity on ambient NO_2 up to 100 kilometers. This is consistent with the results in Clay, Lewis and Severnini (2015). It is for this reason that our wind-and-area-based measure of capacity considers quarter-circles with a radius of 100 kilometers.

4.2 The Effect of Coal-Fired Capacity on IMR

Tables 2 and 3 presents our estimated effects of electricity generating capacity on infant mortality rates (IMR). All of the regressions presented in this subsection are weighted by the number of live births. Panel A of Table 2 estimates the impact of in-district capacity on infant deaths per 1,000 live births overall (Columns 1 and 4), in urban areas (Columns 2 and 5) and rural areas (Columns 3 and 6).[13] Column 1 of this table indicates that a 1GW increase in coal-fired capacity in a district increases infant mortality rates by 14.4% on average. As a point of reference, the average (standard deviation) of in-district coal-fired capacity over our 1996-2014 sample period is 0.14 (0.55).

Our effects are more pronounced in urban areas. Specifically, a 1GW increase in coal-fired capacity increases infant mortality by 19.3% on average. Comparing Columns 2 versus 3, we see that our effect is driven mostly by impacts on urban areas. The effect of coal-fired capacity on infant mortality rates (IMR) in urban areas is positive and statistically significant while we cannot reject the null hypothesis of no effect in rural areas.

Comfortingly, we find no evidence that increases in non-coal capacity impact IMR (Columns 4-6 of Panel A of Table 2). This increases our confidence in our research

[13]We restrict the sample to observations for which urban versus rural breakdowns are available; only 2% of observations are dropped due to this sample restriction.

design because we don't expect electricity production from non-coal-fired sources (i.e.: gas, nuclear, and hydro) to have significant negative health impacts. Moreover, one may think that increased electricity generating capacity results in increased access to electricity and thus health benefits through direct channels such as cooling technologies (e.g. fans) or indirect channels such as increased income. The fact that we find no effect of non-coal capacity on IMR provides suggestive evidence that electricity generating capacity does not, by itself, provide health benefits to the people living in the district where the plant is located relative to other districts in the same state.

[Table 2 about here.]

Panel B of Table 2 presents estimates from our event study framework. We group event time into 5 periods in the interest of conserving statistical power. Event time $\in [-5, -1]$ indicates years that are between 1 and 5 years before the capacity increase; the coefficient estimate associated with this bin should be zero if there are no pre-existing differences between districts that subsequently receive versus don't receive coal-fired capacity. The excluded category is more than 5 years before the event; the estimates presented in Panel B are relative to the effect six or more years before the event.

The results from Panel B indicate that the coefficient on event time $\in [-5, -1]$ is small and statistically indistinguishable from zero for both coal capacity and non-coal capacity. This implies that we fail to reject the null hypothesis of common pre-existing trends: districts that subsequently received exposure to more versus less coal-fired capacity do not exhibit statistically different trends in IMR prior to treatment. Finally, Columns 1 and 2 of panel B indicate that both total IMR and urban IMR spike immediately after new coal capacity is installed, with coal-fired plants having an adverse impact on infant health even more than 10 years after the increase in capacity.

[Table 3 about here.]

To explore the possibility that health impacts from coal-fired capacity spill across district borders, Table 3 considers our wind-and-area-based measure of annual, district-level capacity. For ease of interpretation, we standardize this measure of capacity: a one unit increase in the independent variable corresponds to a one standard deviation increase in this wind-and-area-based capacity measure (1σ). This table demonstrates that coal-fired capacity has a statistically significant and positive impact on IMR in districts downwind from the plant site; this result is stronger in urban areas (Column 2). In contrast, districts upwind from the plant-site see no increase in IMR on average in response to increases in coal-fired capacity (Columns 4-6).

Appendix Table A.7 demonstrates that our effects are not driven by pithead plants (i.e.: plants located near mines), which is important given that coal mining may be associated with air pollution and infant mortality. In addition, we find no evidence of differential changes in overall population density near versus farther away from coal-fired plants before versus after these plants opened; people do not seem to be migrating away in response to the opening of coal-fired power plants in India (see Appendix B.1.1).

4.3 How Large are these Effects?

Our primary finding is that a 1GW increase in coal-fired capacity increases IMR by 15% on average. As a point of comparison, Clay, Lewis and Severnini (2015) finds that a 1GW increase in coal-fired capacity results in a 4.8% increase in IMR for the historical United States.[14] Our estimated effect is roughly 3 times as large as their preferred specification and between 2-3 times as large as their other specifications. There are several possible reasons why the health costs in India are far higher than in the historical United States, including differences in baseline pollution levels, income levels, and characteristics of the production technology (i.e.: plants) and type of coal burned.

We can also assess how the health costs from coal-fired power plants in India compare to the effects from burning coal in other contexts. For example, Beach and Hanlon (2016) estimates the impact of the coal burned by industrial facilities on IMR in 19th century England. In their preferred specification, they find that a one standard deviation increase in coal use increases infant mortality rates by 7.9%. This implies that a 1GW increase in coal-fired capacity in India between 1996-2014 had the same effect on IMR as an almost two standard deviation increase in industrial coal use in 19th century England. Similarly, Cesur, Tekin and Ulker (2017) finds that a one percentage point increase in the residential and commercial heating and cooking done via natural gas rather than coal reduces IMR by 4% in Turkey. Comparing to our estimates, that would imply that shutting down a

[14] We base our comparisons on the estimates from Column 3 of Panel B of Table 4 presented in the working version of the paper posted to NBER website in 2015.

1GW coal fired power plant in contemporary India would have the same effect as a 3.75 percentage point increase in natural-gas-to-coal substitution in Turkey in the 1980s and 1990s. In short, it is difficult to find a historical or contemporaneous analogue to the magnitude of the health costs imposed on the Indian populace by coal-fired power plants.

4.4 Heterogeneity in Effect of Coal-Fired Capacity on IMR

Tables 2 and 3 document robust evidence that increases in coal-fired capacity increase IMR and that this impact is driven almost entirely by urban IMR. There are three possible explanations for why the impacts of coal-fired electricity production are higher for urban areas. First, coal-fired plants tend to be placed closer to urban areas; urban areas within districts may simply be more exposed to coal-fired capacity than rural areas (see Appendix Section B.1). Second, the plants sited closer to urban areas tend to be older than the plants sited in rural areas. Older plants tend to require more input heat to produce the same level of output electricity, resulting in more pollution per unit of capacity. Finally, urban areas tend to have higher baseline pollution levels. If the dose-response function relating pollution to mortality is non-linear, the same increase in coal-fired capacity could have different marginal effects in urban versus rural areas (Hsiang, Oliva and Walker, 2019). Below, we document suggestive evidence supporting both hypotheses.

We find that the effect of coal-fired capacity increases is largest in places with above median baseline levels of NO_2. Specifically, in Appendix Table A.5, we interact annual district-level coal-fired capacity with an indicator that's equal to one if the district's baseline NO_2 level in 1996 is above the median taken over all districts that ever had a coal-fired power plant. This interaction term is positive and statistically significant, suggesting that health costs from increases in coal-fired capacity are especially high in areas with high baseline pollution levels. Finally, Appendix Table A.6 demonstrates that the effect of coal-fired capacity on IMR is far larger for plants built before 2000. This is intuitive: older, less efficient plants must burn more coal to produce the same quantity of electricity. We also show in Appendix Table A.7 that the effect of coal-fired capacity on IMR is larger for plants burning domestic rather than imported coal. Again, this is intuitive: coal imported from Australia or Indonesia typically has lower ash content than coal mined in India.

4.5 Do Plants Provide Local Economic Benefits?

Finally, we ask: does coal-fired capacity convey additional benefits to people living near plant sites relative to the average person living in the state? Since India is characterized by significant institutional barriers and low migration rates, there is no single sufficient statistic such as housing prices that can capture the net benefits of power plants. Instead, we consider how coal-fired capacity impacts a host of different economic outcomes.

First, we consider annual district-level GDP and average luminosity at night. "Night-lights" has been used as a measure of both economic output and rural electrification (Chen and Nordhaus, 2011; Henderson, Storeygard and Weil, 2012; Min et al., 2013; Min and Gaba, 2014; Burlig and Preonas, 2016). We also assess how coal-fired capacity impacts revenues in the manufacturing sector as well as yields in the agricultural sector. Appendix Table A.8 presents the results of this analysis. We find no statistical effect of coal-fired capacity on any of these economic outcomes.

It is important to emphasize that these results do not indicate that coal-fired power plants have no economic benefits. Instead, we simply document that these benefits do not accrue disproportionately to the people living near these plants relative to other people in the same state. This is intuitive given that the primary benefits from power plants is the electricity they produce, and electricity is transported via a transmission grid to places potentially quite far from the plant site. Second, while we consider a number of different economic measures, we can't rule out the absence of any local benefits. Any missing local benefits from coal-fired plants, however, would have to be extremely large to compensate for the massive health costs imposed by these plants on the people nearby.

5 Conclusion and Policy Discussion

Developing countries across the world are building coal-fired power plants to meet rising electricity demand. This paper studies the health costs of the massive expansion of coal-fired electricity production capacity in India over the last two decades. We estimate that a 1 GW increase in coal-fired capacity increased infant mortality rates by roughly 15% relative to other districts in the same state. This magnitude is 2-3 times larger than estimates from the developed world and comparable in magnitude to infant mortality rates associated with deaths due to measles and malaria in India (Million Death Study Collaborators, 2010).

We document three important sources of heterogeneity in the effect of coal-fired capacity on infant mortality rates (IMR). Specifically, the impact of coal-fired capacity on IMR is larger for: (1) older plants, (2) located in districts with higher baseline pollution levels, and (3) burning domestic rather than imported coal. This heterogeneity has at least two immediate policy implications. First, Indian policymakers unveiled a new five-year plan to combat local air pollution in January of 2019 (Abi-Habib and Kumar, 2019). Reducing emissions from coal-fired power plants will undoubtedly be a key part of this (or any) air pollution policy in India. Our work cautions against placing less stringent regulation on older plants in order to get political buy-in from the electricity industry. Indeed, the Clean Air Act (CAA) of 1970 in the United States "grandfathered" older coal-fired plants, decreasing the effectiveness of the CAA as well as increasing the economic costs associated with this policy (Hercher, 1980; Bushnell and Wolfram, 2012). Our second policy implication concerns proposed "protectionist" restrictions on coal imports (Varadhan, 2019). Specifically, our results suggest that these import restrictions may come with substantial environmental costs if coal-fired plants burn domestic coal instead.

75% of the electricity produced in India comes from coal-fired sources and this number might reach as high as 90% by 2030 (Shearer, Fofrich and Davis, 2017). Coal-fired electricity production is an unavoidable reality for the foreseeable future in India. Consequently, the first-order concern for Indian policymakers should be mitigating the health impacts of existing coal-fired plants and ensuring that the siting of new coal-fired plants takes into account their health impacts. Our results indicate that coal-fired plants do not convey disproportionate **local** economic benefits to the people living near these plants; siting and shut-down decisions should thus be based primarily on the environmental health consequences of the plant.

References

Abi-Habib, Maria, and Hari Kumar. 2019. "India Finally Has Plan to Fight Air Pollution. Environmentalists Are Wary." *https://www.nytimes.com/2019/01/11/world/asia/india-air-pollution.html*.

Anderson, Michael L. 2015. "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality." National Bureau of Economic Research.

Arceo, Eva, Rema Hanna, and Paulina Oliva. 2016. "Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City." *The Economic Journal*, 126(591): 257–280.

Auffhammer, Maximilian, Solomon M Hsiang, Wolfram Schlenker, and Adam Sobel. 2013. "Using weather data and climate model output in economic analyses of climate change." *Review of Environmental Economics and Policy*, 7(2): 181–198.

Beach, Brian, and W Walker Hanlon. 2016. "Coal smoke and mortality in an early industrial economy." *The Economic Journal.*

Bechle, Matthew J, Dylan B Millet, and Julian D Marshall. 2013. "Remote sensing of exposure to NO2: Satellite versus ground-based measurement in a large urban area." *Atmospheric Environment*, 69: 345–353.

Bondy, Malvina, Sefi Roth, and Lutz Sager. 2018. "Crime is in the air: The contemporaneous relationship between air pollution and crime."

Burgess, Robin, Olivier Deschenes, Dave Donaldson, and Michael Greenstone. 2014. "The unequal effects of weather and climate change: Evidence from mortality in india." *Cambridge, United States: Massachusetts Institute of Technology, Department of Economics. Manuscript.*

Burlig, Fiona, and Louis Preonas. 2016. "Out of the Darkness and Into the Light? Development Effects of Rural Electrification." Working Paper,,(October).

Bushnell, James B, and Catherine D Wolfram. 2012. "Enforcement of vintage differentiated regulations: The case of new source review." *Journal of Environmental Economics and Management*, 64(2): 137–152.

Cesur, Resul, Erdal Tekin, and Aydogan Ulker. 2017. "Air pollution and infant mortality: evidence from the expansion of natural gas infrastructure." *The Economic Journal*, 127(600): 330–362.

Chen, Xi, and William D Nordhaus. 2011. "Using luminosity data as a proxy for economic statistics." *Proceedings of the National Academy of Sciences*, 108(21): 8589–8594.

Clay, Karen, Joshua Lewis, and Edson Severnini. 2015. "Canary in a Coal Mine: Impact of Mid-20th Century Air Pollution Induced by Coal-Fired Power Generation on Infant Mortality and Property Values." Working paper.

Cropper, Maureen L, Sarath Guttikunda, Puja Jawahar, Zachary Lazri, Kabir Malik, Xiao-Peng Song, and Xinlu Yao. 2019. "Applying Benefit-Cost Analysis to Air Pollution Control in the Indian Power Sector." *Journal of Benefit-Cost Analysis*, 10(S1): 185–205.

Cropper, Maureen, Shama Gamkhar, Kabir Malik, Alexander Limonov, and Ian Partridge. 2012. "The health effects of coal electricity generation in India."

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. "Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings." *The American economic review*, 105(2): 678–709.

Deryugina, Tatyana, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif. 2016. "The mortality and medical costs of air pollution: Evidence from changes in wind direction." National Bureau of Economic Research.

Deschenes, Olivier, Michael Greenstone, and Joseph S Shapiro. 2017. "Defensive investments and the demand for air quality: Evidence from the NOx budget program." *American Economic Review*, 107(10): 2958–89.

Donaldson, Dave, and Adam Storeygard. 2016. "The view from above: Applications of satellite data in economics." *Journal of Economic Perspectives*, 30(4): 171–98.

EPA, S. 1999. "The Benefits and Costs of the Clean Air Act: 1990 to 2010." EPA-410-R99-001. Washington, DC: US Environmental Protection Agency, Office of Air and Radiation.

Gertler, Paul J., Orie Shelef, Catherine D. Wolfram, and Alan Fuchs. 2016. "The Demand for Energy-Using Assets among the World's Rising Middle Classes." *American Economic Review*, 106(6): 1366–1401.

Gibson, Matthew. 2018. "Regulation-induced pollution substitution." *Review of Economics and Statistics*.

Graff Zivin, Joshua, and Matthew Neidell. 2013. "Environment, health, and human capital." *Journal of Economic Literature*, 51(3): 689–730.

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

Greenstone, Michael, and B Kelsey Jack. 2015. "Envirodevonomics: A research agenda for an emerging field." *Journal of Economic Literature*, 53(1): 5–42.

Greenstone, Michael, and Rema Hanna. 2014. "Environmental regulations, air and water pollution, and infant mortality in India." *The American Economic Review*, 104(10): 3038–3072.

Gupta, Aashish, and Dean Spears. 2017. "Health externalities of India's expansion of coal plants: Evidence from a national panel of 40,000 households." *Journal of Environmental Economics and Management*.

Henderson, J Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring economic growth from outer space." *The American Economic Review*, 102(2): 994–1028.

Hercher, David W. 1980. "New Source Performance Standards for Coal-Fired Electric Power Plants." *Ecology LQ*, 8: 748.

Herrnstadt, Evan, and Erich Muehlegger. 2015. "Air Pollution and Criminal Activity: Evidence from Chicago Microdata." National Bureau of Economic Research.

Hsiang, Solomon, Paulina Oliva, and Reed Walker. 2019. "The distribution of environmental damages." *Review of Environmental Economics and Policy*, 13(1): 83–103.

Johnson, Reid, Jacob LaRiviere, and Hendrik Wolff. 2017. "Fracking, coal, and air quality." IZA Discussion Paper 10170.

Lavaine, Emmanuelle, and Matthew Neidell. 2017. "Energy production and health externalities: Evidence from oil refinery strikes in france." *Journal of the Association of Environmental and Resource Economists*, 4(2): 447–477.

Lleras-Muney, Adriana. 2010. "The needs of the army using compulsory relocation in the military to estimate the effect of air pollutants on children's health." *Journal of Human Resources*, 45(3): 549–590.

Luechinger, Simon. 2014. "Air pollution and infant mortality: A natural experiment from power plant desulfurization." *Journal of health economics*, 37: 219–231.

Matheis, Mike. 2016. "Local economic impacts of coal mining in the United States 1870 to 1970." *The Journal of Economic History*, 76(4): 1152–1181.

Million Death Study Collaborators. 2010. "Causes of neonatal and child mortality in India: a nationally representative mortality survey." *The Lancet*, 376(9755): 1853–1860.

Min, Brian, and Kwawu Mensan Gaba. 2014. "Tracking electrification in Vietnam using nighttime lights." *Remote Sensing*, 6(10): 9511–9529.

Min, Brian, Kwawu Mensan Gaba, Ousmane Fall Sarr, and Alassane Agalassou. 2013. "Detection of rural electrification in Africa using DMSP-OLS night lights imagery." *International journal of remote sensing*, 34(22): 8118–8141.

Mittal, Moti L, Chhemendra Sharma, and Richa Singh. 2012. "Estimates of emissions from coal fired thermal power plants in India." 13–16.

Mölter, Anna, Raymond Agius, Frank de Vocht, Sarah Lindley, William Gerrard, Adnan Custovic, and Angela Simpson. 2014. "Effects of long-term exposure to PM10 and NO2 on asthma and wheeze in a prospective birth cohort." *J Epidemiol Community Health*, 68(1): 21–28.

Muller, Nicholas Z. 2014. "Boosting GDP growth by accounting for the environment." *Science*, 345(6199): 873–874.

Muller, Nicholas Z, Robert Mendelsohn, and William Nordhaus. 2011. "Environmental accounting for pollution in the United States economy." *The American Economic Review*, 101(5): 1649–1675.

Munshi, Kaivan, and Mark Rosenzweig. 2009. "Why is mobility in India so low? Social insurance, inequality, and growth." National Bureau of Economic Research.

Sandler, Danielle H, and Ryan Sandler. 2014. "Multiple event studies in public finance and labor economics: A simulation study with applications." *Journal of Economic and Social Measurement*, 39(1, 2): 31–57.

Schlenker, Wolfram, and David B Lobell. 2010. "Robust negative impacts of climate change on African agriculture." *Environmental Research Letters*, 5(1): 014010.

Shearer, Christine, Robert Fofrich, and Steven J Davis. 2017. "Future CO2 emissions and electricity generation from proposed coal-fired power plants in India." *Earth's Future*, 5(4): 408–416.

Tanaka, Shinsuke. 2015. "Environmental regulations on air pollution in China and their impact on infant mortality." *Journal of Health Economics*, 42: 90–103.

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

Varadhan, Sudarshan. 2019. "India's 2018 thermal coal imports grew at fastest pace in four years: sources." *https://www.reuters.com/article/us-india-coal/indias-2018-thermal-coal-imports-grew-at-fastest-pace-in-four-years-sources-idUSKCN1PJ1E1.*

Wolfe, Amir H, and Jonathan A Patz. 2002. "Reactive nitrogen and human health: acute and long-term implications." *Ambio: A journal of the human environment*, 31(2): 120–126.

Wolfram, Catherine, Orie Shelef, and Paul Gertler. 2012. "How Will Energy Demand Develop in the Developing World?" *Journal of Economic Perspectives*, 26(1): 119–38.

Yang, Muzhe, and Shin-Yi Chou. 2017. "The Impact of Environmental Regulation on Fetal Health: Evidence from the Shutdown of a Coal-Fired Power Plant Located Upwind of New Jersey." *Journal of Environmental Economics and Management.*

Table 1: Effect of In-District Capacity on the Log of Pollution

	All Dependent Variables are in Logs							
	Coal				Non-Coal			
	NO_2	$PM_{2.5}$	$PM_{2.5}^{MERRA}$	SO_2^{MERRA}	NO_2	$PM_{2.5}$	$PM_{2.5}^{MERRA}$	SO_2^{MERRA}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Cap (GW)	0.076***	0.015	0.002	0.001	0.033	0.022	0.005	0.009
	(0.018)	(0.016)	(0.004)	(0.012)	(0.026)	(0.014)	(0.006)	(0.008)
R^2	0.968	0.990	0.994	0.994	0.968	0.990	0.994	0.994
Panel B								
$E \in [-5, -1]$	-0.001	0.001	0.001	0.005	0.062	0.005	0.007	0.002
	(0.013)	(0.005)	(0.003)	(0.005)	(0.042)	(0.016)	(0.007)	(0.014)
$E \in [0, 4]$	0.070**	-0.001	-0.001	-0.007	0.076	0.008	0.002	0.006
	(0.029)	(0.013)	(0.003)	(0.011)	(0.046)	(0.022)	(0.011)	(0.015)
$E \in [5, 9]$	0.103**	0.012	-0.002	0.017	0.055	0.022	0.001	-0.006
	(0.038)	(0.031)	(0.008)	(0.023)	(0.040)	(0.027)	(0.014)	(0.022)
R^2	0.968	0.990	0.994	0.994	0.968	0.990	0.994	0.994
Mean of DV	-0.92	2.96	3.71	1.61	-0.92	2.96	3.71	1.61
# of Obs	5,093	4,603	6,884	6,884	5,093	4,603	6,884	6,884
# of Districts	473	469	511	511	473	469	511	511

Notes: Panel A reports the estimated effect of coal-fired electricity production capacity (columns 1-4) and non-coal capacity (columns 5-8) on the log of pollution concentration levels. The pollutants considered are nitrogen dioxide (NO_2) measured by Van Donkelaar et al. (2016) in Columns 1 and 5, fine particulates ($PM_{2.5}$) measured by Van Donkelaar et al. (2016) in Columns 2 and 6, $PM_{2.5}$ measured by the Modern-Era Retrospective analysis for Research and Applications (MERRA) in Columns 3 and 7, and sulfur dioxide (SO_2) measured by MERRA in Columns 4 and 8. Panel B presents results from our event study framework (Sandler and Sandler, 2014). The "event" is an increase in coal-fired capacity and we bin up event time as specified in the table. Event time $\in [-\infty, -6]$ serves as the excluded category and we do not report the "end-point" restriction bin of event times greater than 9 years in the past. All regressions include district fixed effects and state-by-year fixed effects as well as controls for temperature and precipitation. Observations are weighted by district population in 2000. Standard errors are clustered at the state-level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 2: In-District Capacity on the Log Of Infant Mortality Rates

	Coal			Non-Coal		
	Total	Urban	Rural	Total	Urban	Rural
Panel A						
Cap (GW)	0.144*	0.193**	0.093	0.087	0.047	-0.031
	(0.073)	(0.089)	(0.115)	(0.072)	(0.088)	(0.144)
R^2	0.707	0.614	0.738	0.706	0.612	0.738
Panel B						
Event time $\in [-5, -1]$	0.017	0.025	0.050	-0.022	-0.092	0.147
	(0.070)	(0.095)	(0.056)	(0.158)	(0.176)	(0.198)
Event time $\in [0, 4]$	0.065	0.194**	0.069	0.024	0.053	-0.184
	(0.039)	(0.087)	(0.103)	(0.169)	(0.168)	(0.211)
Event time $\in [5, 9]$	0.182	0.205	-0.090	0.094	0.078	-0.110
	(0.184)	(0.199)	(0.142)	(0.165)	(0.149)	(0.177)
R^2	0.708	0.614	0.739	0.706	0.613	0.739
Mean of Dep. Var.	2.029	2.017	1.595	2.029	2.017	1.595
Number of Obs.	6,884	6,884	6,884	6,884	6,884	6,884
Number of Districts	517	517	517	517	517	517

Notes: Panel A of this table reports the effect on log of infant mortality rates (IMR) of total in-district electricity generating capacity summing over coal-fired sources (columns 1-3) versus non-coal-fired sources (columns 4-6). Panel B estimates our event study specification of this effect. In particular, we include separate indicator variables for event time, defined as the difference between the year of observation and year of plant opening/capacity expansion, binned as specified. Event time $\in [-\infty, -6]$ serves as the excluded category and we do not report the "end-point" restriction bin of event times greater than 9 years in the past. All regressions control for temperature and precipitation; we also include district fixed effects and state-by-year fixed effects. Observations are weighted by district-level population in 2000. The top and bottom 1% of IMR has been winsorized. We restrict analysis to observations for which urban and rural breakdowns are nonmissing. Standard errors are clustered at the state-level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

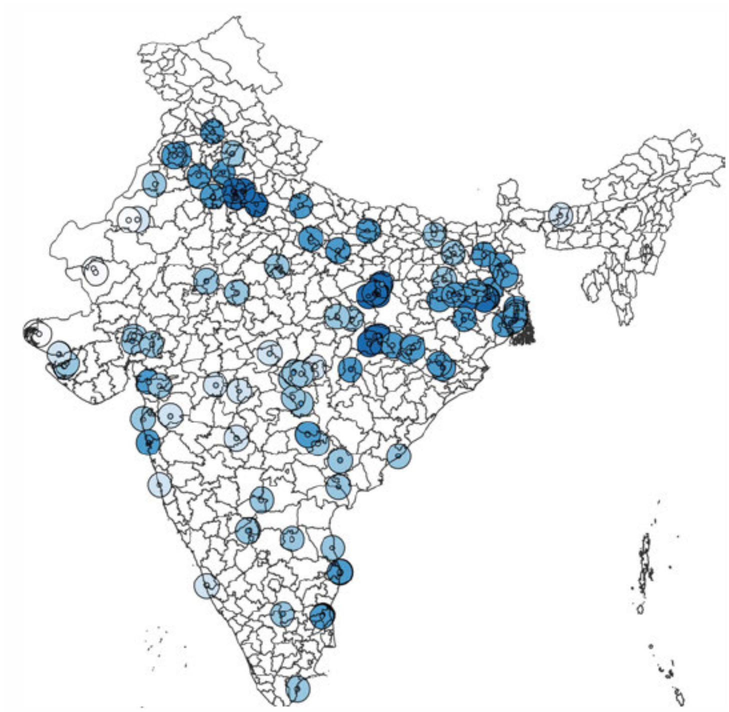
Table 3: Wind-and-Area Based Coal-Fired Capacity on Log of IMR

	Downwind			Upwind		
	Total	Urban	Rural	Total	Urban	Rural
Exposure (1σ)	0.050**	0.080**	0.050	-0.011	-0.015	0.020
	(0.021)	(0.034)	(0.037)	(0.039)	(0.047)	(0.046)
R^2	0.707	0.613	0.738	0.706	0.612	0.738
Mean of Dep. Var.	2.029	2.017	1.595	2.029	2.017	1.595
Number of Obs.	6,884	6,884	6,884	6,884	6,884	6,884
Number of Districts	517	517	517	517	517	517

Notes: Columns 1-3 (4-6) of this table report the effect of our annual district-level measure of area-based coal-fired capacity downwind (upwind) of the plant on the log infant mortality rates (IMR). We compute the downwind (upwind) measure of exposure by calculating the capacity weighted sum over plants of the share of the district's area covered by a quarter-circle with radius 100km pointed downwind (upwind) from the plant; full details are provided in Section 3.1. All regressions control for temperature and precipitation; we also include district fixed effects and state-by-year fixed effects. Observations are weighted by district population in 2000. We winsorize the top and bottom 1% of IMR observations. We restrict analysis to observations for which urban and rural breakdowns are nonmissing. Standard errors are clustered at the state-level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

A Additional Tables and Figures

Figure A.1: Location of Coal-Fired Power Plants with 50km Buffers



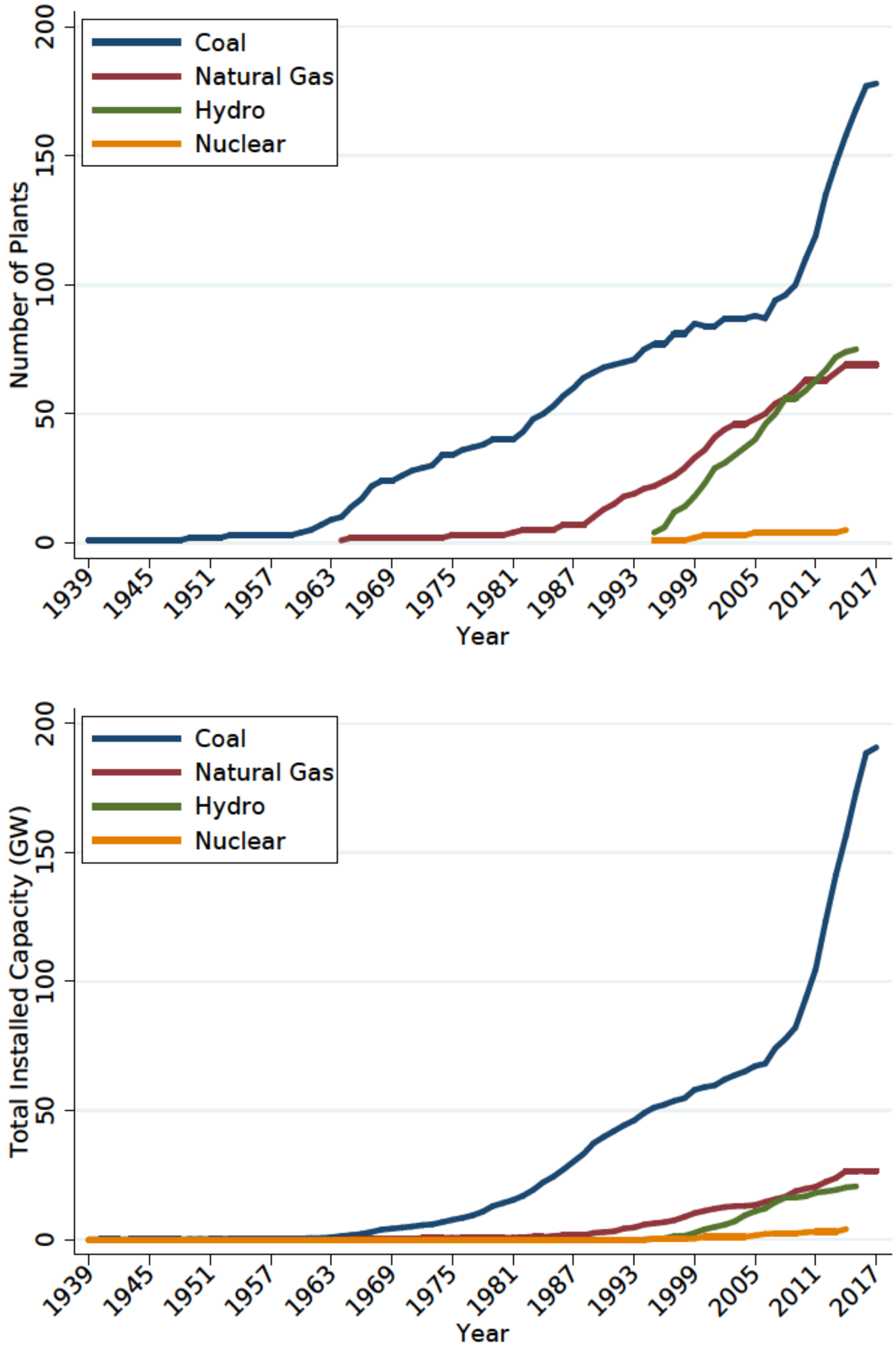
Notes: This figure displays the location of all of the coal-fired power plants in India that operated during any year of our 1996-2014 sample period. We draw a circle with a radius of 50 kilometers around each plant presented in this map.

Table A.1: Summary Statistics: Capacity Measures

	Mean	Std. Dev.
In-District Capacity (GW)		
Coal	0..135	0.551
Noncoal	0.045	0.233
Area-and-Wind Based Capacity (GW)		
$Coal^{Downwind}$	0.178	0.334
$Coal^{Upwind}$	0.184	0.347

Notes: This table presents the summary statistics for total installed capacity and area-based capacity by wind direction. The unit of observation is a district-year, and the sample spans 1996-2014 for all 592 districts.

Figure A.2: Number of Power Plants and Production Capacity over Time By Fuel Type



Notes: The top (bottom) panel plots the total number of plants (total installed electricity generating capacity) of each fuel type in India in each year-of-sample.

Table A.2: Summary Statistics for Outcome Variables and Controls

	Coal Plants *108 Districts*			No Coal Plants *484 Districts*		
	(1) Mean	(2) Std. Dev.	(3) Num. Obs	(4) Mean	(5) Std. Dev.	(6) Num. Obs
Panel A: Outcome Variables						
Infant Mortality Rates						
Total	11.1	10.0	1521	13.2	13.9	6,483
Urban	10.1	11.8	1426	14.1	15.9	6,094
Rural	12.3	11.6	1489	13.5	14.9	5,992
Aggregate Economic Outcomes						
GDP (Bills Rs)	79.5	123.4	924	34.5	42.1	3,645
Night Lights	1.2	3.9	2,177	0.4	1.6	9,359
Manufacturing Outcomes						
Sales (Bills Rs)	48.4	95.7	698	25.3	57.5	2,818
Number of Firms	325.3	524.6	909	184.2	325.3	3,404
Wages (Rs/Day)	146.3	59.4	1,007	141.9	64.8	3,875
Agricultural Outcomes						
Ag Labor (Ths workers)	368.6	270.7	276	252.8	245.3	861
Wage F (Rs/Day)	28.4	24.1	1,332	31.0	28.3	4,063
Wage M (Rs/Day)	42.2	35.8	1,511	43.7	39.9	4,435
Yield (Index)	5.12	0.6	2,119	5.1	0.6	6,869
Pollution						
NO_2 (ppb)	0.58	0.32	1,596	0.38	0.22	6,995
$PM_{2.5}$ ($\mu g/m^3$)	26.15	12.27	1,404	23.21	13.32	6,233
$PM_{2.5}^{M}$ ($\mu g/m^3$)	42.17	19.10	3,996	37.07	19.06	17908
SO_2^{M} ($\mu g/m^3$)	7.04	6.00	3,996	4.52	3.81	17908
Panel B: Control Variables						
Deg Days 21C	1610	451	3,780	1247	671	16,940
Deg Days 29C	194	175	3,780	121	143	16,940
Rainfall	982	530	3,780	1,244	781	16,940

Notes: The unit of observation for this table is district-year. A district is labeled as having "Coal Plants" if had a coal-fired power plant located within its boundardies at any point during our 1996-2014 sample period; all other districts are classified as "No Coal Plants". There are 592 districts in our sample in total. Infant mortality rates are defined as deaths of all children ages 0-1 per 1,000 live births. The top and bottom 1% of IMR has been winzorised for the summary statistics presented in this table. Our data on aggregate gross domestic product (GDP) spans 1999-2012. Manufacturing sales, number of firms, and wages are derived from the Annual Survey of Industries (ASI); our manufacturing data span the sample period 1998-2010. Annual district-level agricultural data are compiled by the International Crops Research Institute for Semi-Arid Tropics (ICRISAT) for the sample period 1979-2014. Finally, the data on average luminosity at night ("night-lights") spans 1993-2013.

Table A.3: Effect of Area-based Coal Capacity by Wind Direction on Log Pollution

	All Dependent Variables are Logged							
	Downwind				Upwind			
	NO_2	$PM_{2.5}$	$PM_{2.5}^M$	SO_2^M	NO_2	$PM_{2.5}$	$PM_{2.5}^M$	SO_2^M
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exposure (1σ)	0.068***	0.001	0.003	0.005	0.042***	0.004	-0.002	-0.004
	(0.013)	(0.004)	(0.002)	(0.004)	(0.010)	(0.006)	(0.003)	(0.005)
R^2	0.969	0.990	0.994	0.994	0.968	0.990	0.994	0.994
Mean of Dep. Var.	-0.92	2.96	3.71	1.61	-0.92	2.96	3.71	1.61
Number of Obs.	5,093	4,603	6,884	6,884	5,093	4,603	6,884	6,884
Number of Districts	473	469	511	511	473	469	511	511

Notes: This table reports the estimated effect of area-based coal-fired electricity production capacity downwind (Columns 1-4) versus upwind (Columns 5-8) from the plant site on the log of pollution concentration levels. We compute the downwind (upwind) measure of exposure by calculating the capacity weighted sum over plants of the share of the district's area covered by a quarter-circle with radius 100km pointed downwind (upwind) from the plant; full details are provided in Section 3.1. The pollutants considered are nitrogen dioxide (NO_2) measured by Van Donkelaar et al. (2016) in Columns 1 and 5, fine particulates ($PM_{2.5}$) measured by Van Donkelaar et al. (2016) in Columns 2 and 6, $PM_{2.5}$ measured by the Modern-Era Retrospective analysis for Research and Applications (MERRA) in Columns 3 and 7, and sulfur dioxide (SO_2) measured by MERRA in Columns 4 and 8. All regressions include district fixed effects and state-by-year fixed effects as well as controls for temperature and precipitation. Observations are weighted by district population in 2000. Standard errors are clustered at the state-level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A.4: Effect of Coal-Fired Capacity on Log of NO_2 by Distance From Plant

Dependent Variable: Log of NO_2 Concentration Levels				
Dist. Radius (in km)	10	50	100	200
Capacity × Close	0.078***	0.060***	0.032**	0.006
	(0.022)	(0.016)	(0.014)	(0.013)
R^2	0.974	0.980	0.980	0.985
Mean of Dep. Var.	-0.782	-0.871	-0.980	-1.042
Number of Obs.	5,962	5,962	5,962	5,962
Number of Plants	180	180	180	180

Notes: This table reports the results from a spatial difference-in-differences model of how coal-fired capacity impacts the log of NO_2 concentration levels at different distances from the plant-site. The unit of observation for the regressions presented in this table are plant/distance buffer/year. Specifically, we stack plant/year observations corresponding to average NO_2 levels within Xkm from the plant and plant/year observations corresponding to average NO_2 levels 500km from the plant; X = 10km, 50km, 100km, or 200km for Columns 1, 2, 3, and 4 of this table respectively. The indicator "Close" is one if the observation corresponds to an average NO_2 level taken based on distance less than 500km. All regressions control for plant/distance-buffer fixed effects and state/year fixed effects, as well as controls for annual temperature and precipitation. Finally, we control for the total capacity of the plant . Standard errors are clustered at the state level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

A.1 Heterogeneous Effects of Coal-Fired Capacity on IMR

Table A.5: In-District Coal-Fired Capacity on IMR: Heterogeneity by NO_2

Dep. Var.: Log of Infant Mortality Rates			
	Total	Urban	Rural
Cap (GW)	-0.055	0.022	-0.169
	(0.094)	(0.096)	(0.219)
Above 50%	0.289**	0.249	0.381
	(0.128)	(0.156)	(0.279)
R^2	0.708	0.614	0.739
Mean of Dep. Var.	2.029	2.017	1.595
Number of Obs.	6,884	6,884	6,884
Number of Districts	517	517	517

Notes: This table reports the effect of annual in-district coal-fired electricity generating capacity on the log of annual district-level infant mortality rates (IMR). In addition to district-level capacity, we include an independent variable constructed by taking the annual district-level sum over plants of annual plant-level capacity times an indicator that's one if the plant is located in an area with NO_2 levels in 1996 greater than the 50% quartile of the distribution of baseline NO_2 levels. The dependent variable considered is the log of IMR; Columns 1, 2, and 3 focus on average IMR across the whole district, across urban areas in the district, and across rural areas in the district respectively. All regressions control for temperature and precipitation; we also include district fixed effects and state-by-year fixed effects. Observations are weighted by district-level population in 2000. We winsorize the top and bottom 1% of IMR and we consider only district/years with nonmissing data for total, urban, and rural IMR. Standard errors are clustered by state. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A.6: In-District Coal-Fired Capacity on Log(IMR): Heterogeneity by Vintage

Dep. Var.: Log of Infant Mortality Rates			
	Total	Urban	Rural
Built Before 2000	0.317*	0.396**	0.177
	(0.175)	(0.187)	(0.182)
Built Between 2000-2010	0.073	0.272	-0.137
	(0.154)	(0.228)	(0.277)
Built After 2010	0.018	-0.007	0.096
	(0.151)	(0.147)	(0.079)
R^2	0.708	0.614	0.739
Mean of Dep. Var.	2.029	2.017	1.595
Number of Obs.	6,884	6,884	6,884
Number of Districts	517	517	517

Notes: This table reports the effect of annual in-district coal-fired electricity generating capacity on the log of annual district-level infant mortality rates (IMR). Each of the three independent variables in these regressions is constructed by taking the district-level sum over plants of annual plant-level capacity times an indicator that's one if the first unit of capacity at the plant site was built: (1) before 2000, (2) between 2000 and 2010, and (3) after 2010 respectively. The dependent variable considered is the log of IMR; Columns 1, 2, and 3 focus on average IMR across the whole district, across urban areas in the district, and across rural areas in the district respectively. All regressions control for temperature and precipitation; we also include district fixed effects and state-by-year fixed effects. Observations are weighted by district-level population in 2000. The top and bottom 1% of IMR has been winsorized and we consider only district/years with nonmissing data for total, urban, and rural IMR. Standard errors are clustered by state. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A.7: In-District Coal-Fired Capacity on IMR: Pithead versus Imported

Dep. Var.: Log of Infant Mortality Rates	Total	Urban	Rural
Cap: Any (GW)	0.288***	0.403***	0.043
	(0.100)	(0.132)	(0.139)
Pithead Cap	-0.235	-0.370*	0.489
	(0.171)	(0.206)	(0.420)
Import Cap	-0.244	-0.330**	-0.175
	(0.151)	(0.143)	(0.221)
R^2	0.706	0.614	0.738
Mean of Dep. Var.	2.026	2.017	1.598
Number of Obs.	6,679	6,679	6,679
Number of Districts	503	503	503

Notes: This table reports the effect of annual in-district electricity generating capacity on the log of annual district-level infant mortality rates (IMR). In addition to district-level capacity, each of the other two independent variables is constructed by taking the annual district-level sum over plants of annual plant-level capacity times an indicator that's one if: (1) the plant's modal coal delivery is classified as coming from a nearby ("pithead") mine (labelled "Pithead Cap") and (2) if more than 10% of the overall quantity of coal delivered is classified as "imported" (labelled "Import Cap") . The dependent variable considered is the log of IMR; Columns 1, 2, and 3 focus on average IMR across the whole district, across urban areas in the district, and across rural areas in the district respectively. All regressions control for temperature and precipitation; we also include district fixed effects and state-by-year fixed effects. Observations are weighted by district-level population in 2000. The top and bottom 1% of IMR has been winsorized and we consider only district/years with nonmissing data for total, urban, and rural IMR. Standard errors are clustered by state. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A.8: Effect of In-District Capacity on Economic Outcomes

	Coal				Non-Coal			
	log(GDP)	NL	log(Rev)	log(Yield)	log(GDP)	NL	log(Rev)	log(Yield)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Cap (MW)	-0.020	-0.015	0.079	-0.013	0.009	-0.006	0.158	-0.026
	(0.047)	(0.087)	(0.257)	(0.024)	(0.012)	(0.135)	(0.175)	(0.084)
R^2	0.993	0.910	0.924	0.924	0.993	0.910	0.924	0.924
Panel B								
$E \in [-5, -1]$	0.004	0.052	-0.079	0.004	-0.027	0.153**	-0.073	0.025
	(0.011)	(0.084)	(0.048)	(0.017)	(0.022)	(0.072)	(0.090)	(0.064)
$E \in [0, 4]$	-0.002	-0.050	0.087	-0.003	0.001	0.108	0.062	-0.029
	(0.039)	(0.103)	(0.216)	(0.030)	(0.032)	(0.117)	(0.164)	(0.071)
$E \in [5, 9]$	-0.056	-0.149	0.091	-0.015	-0.017	-0.150	0.227	0.063
	(0.043)	(0.144)	(0.391)	(0.029)	(0.037)	(0.188)	(0.214)	(0.110)
R^2	0.993	0.910	0.924	0.924	0.993	0.910	0.924	0.924
Mean of Dep. Var.	3.826	0.684	1.833	5.161	3.826	0.684	1.833	5.161
Number of Obs.	4,569	11,495	5,406	8,956	4,569	11,495	5,406	8,956
Number of Districts	487	558	521	295	487	558	521	295

Notes: Panel A of this table reports the effect of total in-district electricity generating capacity summing over coal-fired sources (columns 1-4) versus non-coal-fired sources (columns 5-8) on a host of different economic outcomes. Specifically, the dependent variable considered in Columns 1 and 5 is the log of annual district-level gross domestic product (GDP); we have GDP data for the sample period 1999-2012. The dependent variable for Columns 2 and 6 of this table is the average luminosity at night ("night-lights"); the regressions in these two columns span the sample period 1993-2013. Columns 3 and 7 present effects of capacity on the log of annual district-level total manufacturing revenues; data on manufacturing revenues comes from the Annual Survey of Industries (ASI) for the sample period 1998-2010. Finally, the dependent variable considered in Columns 4 and 8 is the log of agricultural yields, which is compiled by the International Crops Research Institute for Semi-Arid Tropics (ICRISAT). This data-set spans the sample period 1979-2014. Panel B estimates our event study specification (Sandler and Sandler, 2014). In particular, we include separate indicator variables for event time, defined as the difference between the year of observation and year of plant opening/capacity expansion, binned as specified. Event time $\in [-\infty, -6]$ serves as the excluded category and we do not report the "end-point" restriction bin of event times greater than 9 years in the past. All regressions control for temperature and precipitation; we also include district fixed effects and state-by-year fixed effects. Observations are weighted by district-level population in 2000. Standard errors are clustered at the state-level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

B Population Density and Geographic Sorting

B.1 Effect of Coal-Fired Capacity on Population Density

This Appendix section investigates whether coal-fired power plants are more likely to be sited in urban areas versus rural areas. If coal-fired plants tend to be placed near urban areas, this could explain why coal-fired capacity impacts IMR in urban areas but not rural areas.

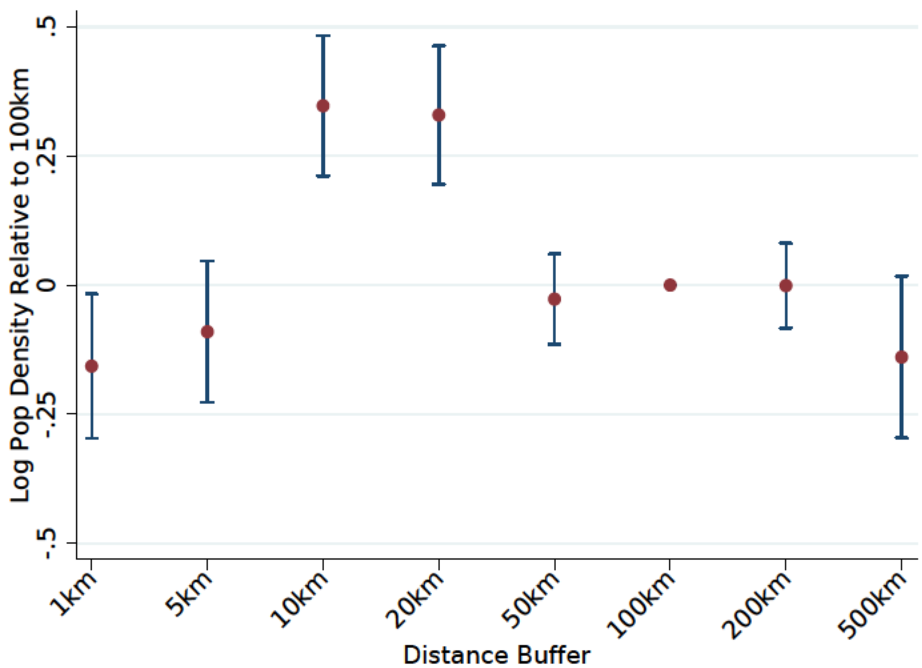
To this end, we first calculate population density around each coal plant site. Specifically, for each of our 180 power plants, we compute average population density within 1km of the plant site as well as between 1km-5km, 5km-10km, 10km-20km, 50km-100km, 100km-200km, and 200km-500km from the plant site. We use data from . If it is true that coal-fired plants are placed closer to urban areas, then population densities should be higher closer to plant sites relative to farther away from these sites. To test this, we estimate:

$$log(D_{p,t,b}) = \alpha_{p,t} + \sum_{l=1}^{B} \beta^{l} * 1(b = l) + \epsilon_{p,t,b} \tag{5}$$

where $log(D_{p,t,b})$ is the log of population density in year t in a "donut" who's outer ring has radius b around plant p. We include plant-year fixed effects (i.e.: $\alpha_{p,t}$). Standard errors are clustered by district.

The coefficient estimates, along with 95% confidence intervals, from estimating Equation (5) are presented in Figure B.1. We omit the dummy variable for the 50km-100km buffer, so all point estimates are relative to a distance of 100km. In Figure 5, we find that population density is nonlinear in distance from the plant. The density within 5km of the plant is no greater than the density more than 50km from the plant site. However, population density drops when moving from 10km-20km to 20km-50km. This suggests that coal-fired plants tend to be placed roughly 10km-20km from areas with high population density (i.e.: urban areas)

Figure B.1: Average Population Density at Different Distances from Plants



Notes: This figure reports point estimates and 95% confidence intervals from estimating Equation (5). Specifically, for each of our 180 coal-fired power plants, we compute average population density within 1km of the plant site as well as between 1km-5km, 5km-10km, 10km-20km, 50km-100km, 100km-200km, and 200km-500km from the plant site. These population data come from the Socioeconomic Data and Applications Center (SEDAC), a data center in NASA's Earth Observing System Data and Information System (EOSDIS). We stack the average population density corresponding to each of these distance buffers, giving us a plant/buffer/year level data-set. We then regress an indicator for distance buffer on average population density, controlling for plant/buffer fixed effects. The area between 50km to 100km serves as the excluded category, so point estimates are relative to density in this area. Standard errors are clustered by district.

B.1.1 Geographic Sorting

In this subsection, we test for aggregate sorting in response to changes in coal-fired capacity. We estimate two specifications:

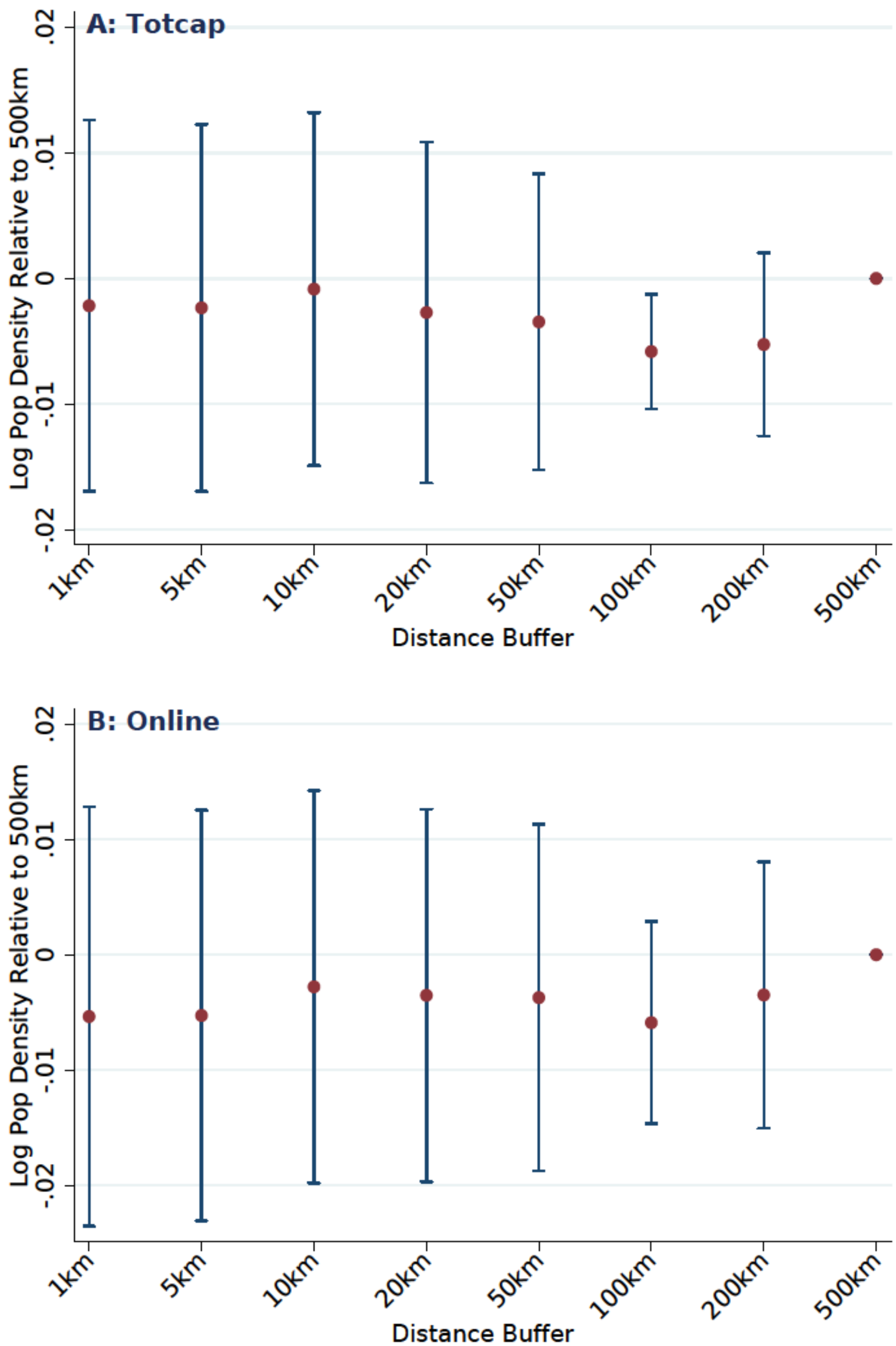
$$log(D_{p,t,b}) = \alpha_{p,b} + \theta_{s,t} + 1(\text{Opened})_{p,t}(\beta_0 + \beta_1 1(b = \text{Not 500km Buffer})) + \epsilon_{p,t,b} \quad (6)$$

$$log(D_{p,t,b}) = \alpha_{p,b} + \theta_{s,t} + Cap_{p,t}(\beta_0 + \beta_1 1(b = \text{Not 500km Buffer})) + \epsilon_{p,t,b} \quad (7)$$

where $log(D_{p,t,b})$ is the log of population density in year t in a circle of radius b around plant p. We include plant/distance-buffer fixed effects $\alpha_{p,b}$ as well as state/year fixed effects $\theta_{s,t}$. The coefficient of interest is β_1, the interaction effect from increases in coal-fired capacity (or plant openings) on log population density closer rather than farther from the plant. If a plant coming online or increases in capacity lead to outward migration, we would expect $\beta_1 < 0$: less population density at closer distances to the plant relative to 500km from the plant after the plant comes online (or the plant increases capacity). Standard errors are clustered by district.

Point estimates and 95% confidence intervals corresponding to Equation 6 (Equation 7) are presented in the top panel (bottom panel) of Figure B.2. For both specifications, we see precisely estimated zeros at all distance bandwidths. Population density does not appear to change after plants come online (or increase capacity) near relative to far from the plant. In short, we find no evidence of sorting/out-migration in response to either plants coming online or capacity increases at these plants.

Figure B.2: Average Population Density at Different Distances from Plants



Notes: This figure reports point estimates and 95% confidence intervals corresponding to Equation 6 (Equation 7). Specifically, for each of our 180 coal-fired power plants, we compute average population density within 1km of the plant site as well as between 1km-5km, 5km-10km, 10km-20km, 50km-100km, 100km-200km, and 200km-500km from the plant site. These population data come from the Socioeconomic Data and Applications Center (SEDAC), a data center in NASA's Earth Observing System Data and Information System (EOSDIS). We stack the average population density corresponding to each of these distance buffers, giving us a plant/buffer/year level data-set. We then regress an indicator for whether the plant is online in the year (top panel) or the capacity of the plant in the year (bottom panel) on average population density in each distance ring. The area between 200km to 500km serves as the excluded category, so point estimates are relative to density in this area. We include plant/distance-buffer fixed effects as well as state/year fixed effects. Standard errors are clustered by district.