The Long-Run Effects of Coal-Fired Power Plants on Neonatal, Infant, and Child Health in India

[Preliminary Draft - Do not Cite]

Raahil Madhok[1], Rohini Pande[2], Kevin Rowe[3], and Anish Sugathan[4]

[1]University of British Columbia
[2]Yale University
[3]Uber
[4]Indian Institute of Management, Ahmedabad

October 27, 2022

Abstract

This paper quantifies the long-run effects of India's dramatic expansion of coal-fired electric power plants from 1970 to 2017 on neonatal, infant, and child health. We use a particle trajectory model that draws wind-driven patterns of emissions dispersion from power plants to construct an exogenous measure of cumulative pollution exposure over three decades. Our measure of historic power plant exposure predicts present-day air quality three to six times better than conventional wind-direction instruments in the literature. A one standard deviation increase in long-run exposure to power plant emissions increases neonatal, infant, and child deaths by 0.6 (2 percent of the mean), 0.9 (1.9 percent), and 1.3 (2.2 percent) per 1000 live births, respectively. These effects are largely driven by exposure in-utero, as well as exposure to private power plant clusters forming between 1992-2005. We find no evidence of differential economic development between more- and less-exposed districts, ruling out adaptation, and underscoring pollution and the main mechanism.

1 Introduction

Increased electricity demand is a characteristic feature of economic development. One-third of global electricity is generated by coal-fired power plants, with China and India accounting for 85% of new coal power capacity since 2005 (Shearer et al., 2019). While electricity access is critical for economic progress, coal combustion also releases harmful air pollutants. In 2015 alone, outdoor air pollution resulted in about 4.5 million deaths globally (Landrigan et al., 2018; Murray et al., 2020). Nearly 1 million of these deaths occurred in India, the most of any country.

While an emerging literature quantifies short-run impacts of exposure to poor air quality, less is known about long-term impacts largely due to difficulty identifying quasi-random variation in long-run air quality across geographic units. Yet, a deeper understanding is necessary given that many of the largest sources of air pollution are long-lived infrastructure assets, like power plants and highways. Thus, estimates of the long-run effects of differences in ambient air quality are arguably more important to policy than the short-run effects which have been studied extensively in the literature.

The main contribution of this paper is to document the impact of long-run exposure to coal-fired power plant pollution on early-life mortality in India from 2000 to 2017. India is an ideal setting to study this issue because power plants constitute the largest point source of outdoor air pollution in India. Despite the rapid growth of solar investment in recent years, India still relies on coal-fired power to supply two-thirds of its electricity. Even if growth rates of renewables are sustained, and no new coal power plants are built, coal is projected to remain India's majority energy source for at least ten more years, and the dominant source until 2040 (International Energy Agency, 2021). This inertia reflects over five decades of concerted investment in coal-fired power, including more than a doubling of installed capacity since 2000.

The closest point of comparison is from two short-run studies. Most related, Barrows et al. (2019) find that a 1 standard deviation increase in coal-fired power plant capacity in Indian districts results in 15% higher infant mortality. In a developed country setting, the equivalent mortality impact is 6.5% in mid-20th century United States counties (Clay et al., 2016). Both studies leverage annual panel data over several decades, which is suited for identifying short-run, but not long-run, impacts. One reason that extrapolating these estimates to the long-run is problematic is because effects of exposure may accumulate over time, such that the cumulative effect of long-term exposure exceeds the sum of short-term coefficients. Instead, what is needed is a direct source of exogenous variation in long-run exposure to coal power plant pollution.

We overcome this challenge by first compiling detailed data on the location of all coal power plants constructed in India between 1970-2017. We then compute exposure to emissions from these plants using a state-of-the-art atmospheric dispersion model. The model simulates a "puff" of tracer particles emanating from each chimney, draws their three dimensional wind patterns

(known as wind fields) over space and time, and shows where the particles land on a 10×10km grid. We run the model on a plant-by-year basis and then compute the cumulative exposure at the district level over thirty years. This yields a novel, plausibly exogenous, measure of a district's total population exposure to upwind power plant pollution over three decades.

Our model-generated exposure measure is more predictive of ambient pollution than standard wind-direction measures from the literature. First-stage regressions of Sulfur Dioxide (SO_2) and Nitrogen Dioxide (NO_2) concentration on historic power plant exposure yield F-statistics between three and six times greater for our measure compared to the conventional measure. Our approach also has the advantage of imposing no restrictions on the distance over which emissions travel, allowing us to study health outcomes manifesting hundreds of kilometers from power plants.

To estimate long-term health impacts, we combine historic power plant exposure with new district-level data on neonatal, infant, and child mortality compiled by Dandona et al. (2020) from a variety of demographic surveys and vital registration metrics. These data are available for the years 2000, 2007, and 2017, permitting a research design based on three separate cross-district comparisons within states. By exploiting cross-sectional variation, we can identify plausible control groups for historically exposed districts: other districts in the same state, located at equal distance to power plants, but exposed to less power plant pollution in the past three decades due to differences in wind fields.

Our analysis yields three key findings. First, children in districts consistently exposed to coal-fired power plants over the past three decades experience significantly more deaths during the first five years of life compared to cleaner districts in the same state. A one standard deviation increase in historic exposure causes 0.6 more neonatal deaths (2% of the mean), 0.9 more infant deaths (1.9%), and 1.3 more child deaths per 1000 live births.

Second, worse ambient air quality in exposed districts is the main mechanism driving our mortality estimates. We find that ambient SO_2 and NO_2 concentrations are 50% and 16% higher, respectively, in districts historically more exposed to power plants. Although air quality may not be the *only* channel through which power plant exposure affects health, we underscore its importance by ruling out two competing mechanisms. We rule out differential adaptation by showing that there is no difference in present-day measures of economic development—including literacy, poverty rates, and nightlight intensity—between historically exposed and non-exposed districts in the same state. We also show no evidence of different migrant shares, helping rule out endogenous sorting as a potential explanation for our results.

Third, emissions exposure in-utero as well as during the 1992-2005 period are the critical periods influencing present-day mortality. We arrive at this result by decomposing cumulative exposure into distinct time periods marked by policy-driven growth in electricity generation. 1992-2005 is the period after which private players were allowed to enter the power generation sector in India. Our results imply that present-day mortality from historic emissions is mainly influenced by

plant clusters forming after the 1991 deregulation.

Contributions to the Literature. The main contribution of this paper is to credibly estimate the health impacts of long-run exposure to coal-fired power plants, an important gap in the literature. A deeper understanding of long-run impacts is important since most pollution regulations strive for long-term improvements in air quality.

Most related literature documents *short-term* health effects of air pollution with panel data in developed countries (Chay and Greenstone, 2003; Currie and Neidell, 2005; Currie and Walker, 2011; Knittel et al., 2016; Deryugina et al., 2019), and to a lesser extent in developing countries (Arceo et al., 2016; Jayachandran, 2009; Barrows et al., 2019). Among the few papers documenting long-term effects, almost all elicit later-life health impacts of early-life pollution exposure (Anderson, 2019; Rosales-Rueda and Triyana, 2019; Currie and Vogl, 2013). Besides the few studies on the impacts of in-utero exposure on early-life outcomes (Currie and Neidell, 2005; Von der Goltz and Barnwal, 2019), we are unaware of studies covering a longer historic period of exposure.

The key to answering our research question is identifying a source of exogenous variation in long-run exposure to coal power plants. The cornerstone of our identification strategy is an atmospheric dispersion model that simulates the trajectory of tracer particles from any point source based on 3D wind fields and other meteorological characteristics. Many previous studies instead rely on constructing treatment (exposed) and control (non-exposed) groups based on distance from a polluting source (Currie et al., 2015; Jayachandran, 2009; Komisarow and Pakhtigian, 2022). Some leverage variation in wind direction for additional exogeneity (Deryugina et al., 2019; Barrows et al., 2019; Herrnstadt et al., 2021). Both approaches require restrictive assumptions about the area and distance of pollutant dispersal, which generally results in limiting the treatment area to a radius around the emissions source that may reflect only a small fraction of the distance over which emissions disperse. Moreover, wind direction is better suited for predicting seasonal variation in pollution but loses substantial explanatory power when aggregated for long-term analyses. In contrast, our model-generated measure requires no assumptions about plume shape or size and retains all spatial variation during aggregation. It also allows us to capture a greater share of the total air pollution exposure due to emissions.

Our measurement approach is conceptually similar to previous studies that identify quasi-random variation in exposure to environmental change. For example, Duflo and Pande (2007) study the impact of dams in India. Since dam placement is endogenous, they use differences in river gradient to identify plausibly exogenous variation. Lipscomb et al. (2013) and Blakeslee et al. (2020) use a similar approach to measure exposure to hydroelectricity and groundwater shortages in Brazil and India, respectively. Whereas we also rely on an underlying source of quasi-random variation (the dispersion model), our approach differs since we are interested in impacts of exposure to emissions from power plants, not the impact of the plants per se. As such, districts without

power plants, and possibly quite far from them, may be treated in our analysis.

Methodologically, we join the literature leveraging cross-sectional research designs to elicit long-term impacts of environmental change. First introduced by Mendelsohn et al. (1994), this approach relies on comparing outcomes across geographic units with different long-term climate averages. Identification relies on inclusion of many geography covariates and a selection-on-observable assumption, which is prone to omitted variable bias (Deschênes and Greenstone, 2007). Druckenmiller and Hsiang (2018) introduce a new estimator, called spatial first differences, which allows credible inference even with omitted variables. However, the approach only makes comparisons between neighbors, which is too restrictive for our application. Since our measure of historic exposure is plausibly orthogonal to geography, our cross-sectional approach is more flexible and less reliant on selection-on-observables.

In addition to providing new evidence on the impacts of historic pollution exposure on early-life mortality, we also advance a smaller literature on the impacts of exposure to power plants specifically. To our knowledge no other study has estimated the long-run health impacts of sustained exposure to coal-fired power plants. Heblich et al. (2016) document path-dependence in UK neighborhood composition following historic power plant investments, but do not investigate health disparities. In the short-run, Barrows et al. (2019), Cropper et al. (2021), and Gupta and Spears (2017) find increases in infant mortality, total mortality, and coughing incidence, respectively, from power plant exposure in India. Our results imply that adverse health outcomes in the short-run are also observable in the long-run, implying minimal adaptation. This contribution is especially important given that coal is expected to be India's dominant fuel source for the next several decades.

The rest of this paper is organized as follows. The next section describes the data and, in particular, details of the particle trajectory model. Section 3 presents three motivating facts gleaned from the data along with our main estimating equations. Sections 4 and 5 presents the main results and robustness checks, respectively.

2 Data Construction

We use three main data sources to study the health effects of long-run exposure to coal power plants: the universe of thermal power plants operating from 1970-2017; data on neonatal, infant, and child mortality in 2000, 2007, and 2017 for all districts in India; and high-resolution satellite data on ambient air pollution for the same three years. The key identification challenge is finding exogenous variation in long-run exposure to power plants. We employ a particle trajectory model that uses wind-fields to trace particle paths from each smoke stack in our sample. This section describes our data and how we measure long-term power plant exposure.

2.1 Main Variables

Thermal Power Plants. Thermal coal-fired power plants generate the overwhelming majority of India's electricity. We obtained annual installed capacity reports from all utility-scale plants operating in India between 1970 and 2017 from the Central Electricity Authority (CEA). There were 217 coal power stations with 757 generating units in operation during these 47 years. We manually geolocate each plant using multiple sources and then verify its presence using visual inspection of satellite imagery. Next, we use the plant coordinates to identify the district of construction and aggregate capacity to the district-annual level. This produces a record of thermal coal-fired capacity expansion within districts over five decades for the whole country.

Ambient Air Pollution. We use high-resolution satellite-based measures of SO_2, NO_2, and particulate matter ($PM_{2.5}$). Gridded SO_2 data is retrieved in separate monthly files for 2004-2017 from the Ozone Monitoring Instrument (OMI) aboard NASA's Aura satellite at 0.25×0.25 degree resolution. Cell values represent SO_2 concentration in the lower troposphere in Dobson Units (2.687×10^{-20} molecules$/m^2$). Gridded NO_2 data for the same period is also retrieved from the OMI satellite at 0.25×0.25 degree resolution. Cell values are measured in molecules$/cm^2$. Both data products undergo substantial cleaning, validation, and outlier detection by NASA before public release.

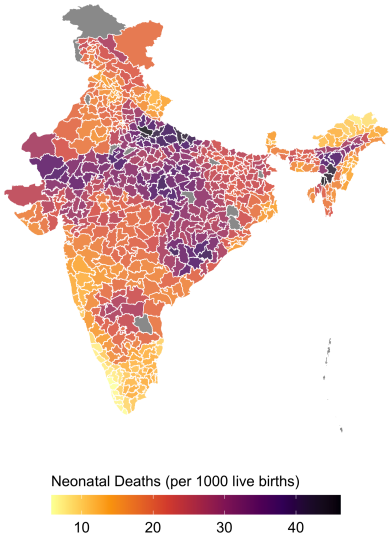
$PM_{2.5}$ is an amalgamation of multiple underlying particulate species and not measured directly by any satellite instrument. We instead use a reanalysis data product developed by Van Donkelaar et al. (2016). This product provides estimates of $PM_{2.5}$ concentrations by assimilating concentrations of underlying particles (e.g. black carbon, organic carbon) through a geo-chemical transport model to produce ground-level $PM_{2.5}$ estimates. Data is provided annually at 0.1×0.1 degree resolution for 1996-2017.

We construct a district-annual pollution panel in two steps. First, we extract weighted mean pollution across cells within district boundaries as delineated by the 2011 Census. Cells overlapping multiple districts contribute to each district mean in proportion to their overlap fraction. Second, we aggregate over months to the district-annual mean (for SO_2 and NO_2).

Neo-Natal, Infant, and Child Mortality. We obtain mortality data at the district level for the years 2000, 2007, and 2017 calculated as part of the 2017 Global Burden of Disease (GBD) study[1]. To our knowledge, these are the most comprehensive and recently validated sub-national mortality estimates for India. District mortality is measured as number of neo-natal (0-28 days), infant (< 1 year), and child (< 5 years) deaths per 1000 live births. Estimates are assembled from numerous underlying data sources according to a standardized procedure used for all 195 studies included in the GBD. For India, the main data sources include: the Sample Registration System, Vital Registra-

[1] Data is available from Dandona et al. (2020)

Figure 1: District-wise Neonatal Mortality (2017)



Note: Colours describe neonatal deaths per 1000 live births. Data from Dandona et al. (2020). Grey indicates NA.

tion Systems, Population Censuses, the National Family Health Surveys, District Level Household Surveys, and Annual Health Surveys. Details of all data sources and calculations are provided in Dandona et al. (2020).

Data are provided for 2017 district delineations (723 districts). We manually map these back to their "parent" districts as per 2011 borders for consistent matching across datasets. Figure 1 illustrates the spatial distribution of neonatal mortality across districts in 2017.

Determinants of Plant Placement. We also collected data on a range of determinants of power plant placement to include as covariates in our analysis. Power plants are typically sited in districts with water bodies, coal deposits, and low elevation (see section 3.1 for details). These districts may experience higher pollution exposure by virtue of proximity, and may also develop differently than districts far away from power plants.

We select plant placement covariates based on engineering considerations and government guidelines (Authority, 2012; Central Electricity Authority, 2010; Mays et al., 2011; Planning Commission, 1961). Digital vector maps of coal deposits are obtained from the United States Geological Survey (Trippi and Tewalt, 2011). Rivers, lakes, and reservoir shapefiles are obtained at 10m resolution from Natural Earth Data.[2] A gridded digital elevation map is obtained from the NOAA at

[2]Accessed from: http://www.naturalearthdata.com/downloads/10m-physical-vectors/

10km $\times$ 10km resolution.[3]

We compute numerous plant placement covariates from these digital maps, including: distance from district centroid to nearest power plant, distance to nearest coal deposit, area of nearest deposit, distance to water bodies, area of nearest water body, elevation, and slope.

Adaptation. We collect data on a wide range of development outcomes to test for differential adaptation between districts exposed and not exposed to power plants. The majority of development indicators are obtained from the Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG) (Asher et al., 2021). The main variables include: literate population share, percentage of households below the 2012 national poverty rate (27.2 Rs./day in villages and 33.3 Rs./day in towns), per capita consumption (Rs.), and total employment in the service sectors. All variables are provided at the village/town level and then aggregated up to the district.[4]

Endogenous sorting is another adaptation response that poses an identification challenge in long-term studies (Kuminoff et al., 2013). We test for this using recently released migration data from the 2011 Census. The Census D3 tables report the district migrant population by type (i.e. for marriage, employment, with whole family) and by when migrants arrived (i.e. 1-4 years ago, 5-9 years ago, etc.). We use the number of migrants who in-migrated with their whole family to study endogenous sorting in section 5.2.

2.2 The HYSPLIT Particle Trajectory Model

Our empirical strategy relies on plausibly exogenous long-run variation in exposure to air pollution emissions from power plants. The distinguishing feature of our approach is an atmospheric dispersion model that draws wind-driven trajectories of tracer particles from each plant chimney. We use this to derive estimates of how exposed districts have been to power plant pollution over the past three decades. This section describes how long-run exposure is calculated and then discusses the advantages of our modelling approach.

Dispersion Model Overview. District exposure to power plant emissions in the long-run is measured using the the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model developed by the NOAA (Draxler and Hess, 1998; Draxler et al., 2020). HYSPLIT is a modeling system for computing simplified dispersion trajectories widely used by the atmospheric science community. Although originally designed for rapid response to atmospheric emergencies, such as tracking downstream exposure to toxic gas leakage and volcanic ash (Stein et al., 2015), HYSPLIT

[3] Accessed from: https://www.ngdc.noaa.gov/mgg/topo/globe.html

[4] Variables in the SHRUG are assembled from different administrative sources and mapped to consistent geographic units. The underlying data sources for our variables include the 2011 Census (literacy), 2012 Socio-economic and caste census (poverty, consumption), and 2013 Economic census (employment).

can generate dispersion plumes from any point source. The model uses meteorological wind fields to determine the 3D path of computational air parcels emanating from large point sources–in our case, power plants—to estimate the spatial distribution of wind-driven exposure.

We use HYSPLIT because of its simplicity. More advanced chemical transport models (CTMs) have estimated spatiotemporal trajectories of downstream air pollution based on complex atmospheric chemistry and physical characterization of multiple particle species at high resolution (Guttikunda and Jawahar, 2014). The computational complexity of these models make it impractical for studying downstream exposure profiles of hundreds of point sources over several decades. In contrast, HYSPLIT produces simplified emission dispersion plumes almost entirely by pre-generated 3D wind fields. This simplifies full-scale CTMs in exchange for computational scalability (Henneman et al., 2019, 2021).

Long-term Exposure Computation. We use version 4 of HYSPLIT to measure long-term exposure to coal power plant emissions in India. Our approach mirrors Henneman et al. (2019), who model coal power plant exposure in the United States at the zip-code level. Their low complexity method predicts actual downstream pollution levels with high accuracy, comparable to more complex CTMs (Henneman et al., 2021).

We calculate long term exposure in three steps. First, we draw bi-monthly exposure plumes for each power station from the date of commissioning until 2017. Figure 2A show the location and capacity of all stations in our sample. To construct plumes, we first generate a 0.1×0.1 degree grid covering all of India and then run the trajectory model from the plant coordinates for 24 hours every 14 days. The model simulates a computational "parcel" of air, represented by the model default fixed number of mass-less particles emitted from the smokestack. The count of particles landing in each grid cell, within a fixed atmospheric boundary layer height of 900m, at the end of the 24-hour run is used as an estimate of the fraction of emission emanating from the corresponding source smokestack. We use a parallel computing framework to reduce the computation burden and produce separate time-varying dispersion plumes for each of the 217 power plants in our sample. The result is a series of plant-level exposure rasters describing the downstream concentration of mass-less computational air parcels $EmissionCount_{ijt}$, at each receptor grid cell i, emitted from coal power station j at time t.

Second, we sum over plant-level rasters in each year to calculate the total exposure of grid cell i to all coal power plants operating in a given year, adjusting for the fact that exposure to a mega power plant is different than a small plant. Specifically, downstream emission exposure over the full grid, $GridExposure_{it}$, is calculated by weighting the normalized, gridded $EmissionCount_{ijt}$ concentration distribution of each coal station j by its respective generating capacity $Capacity_{jt}$ in

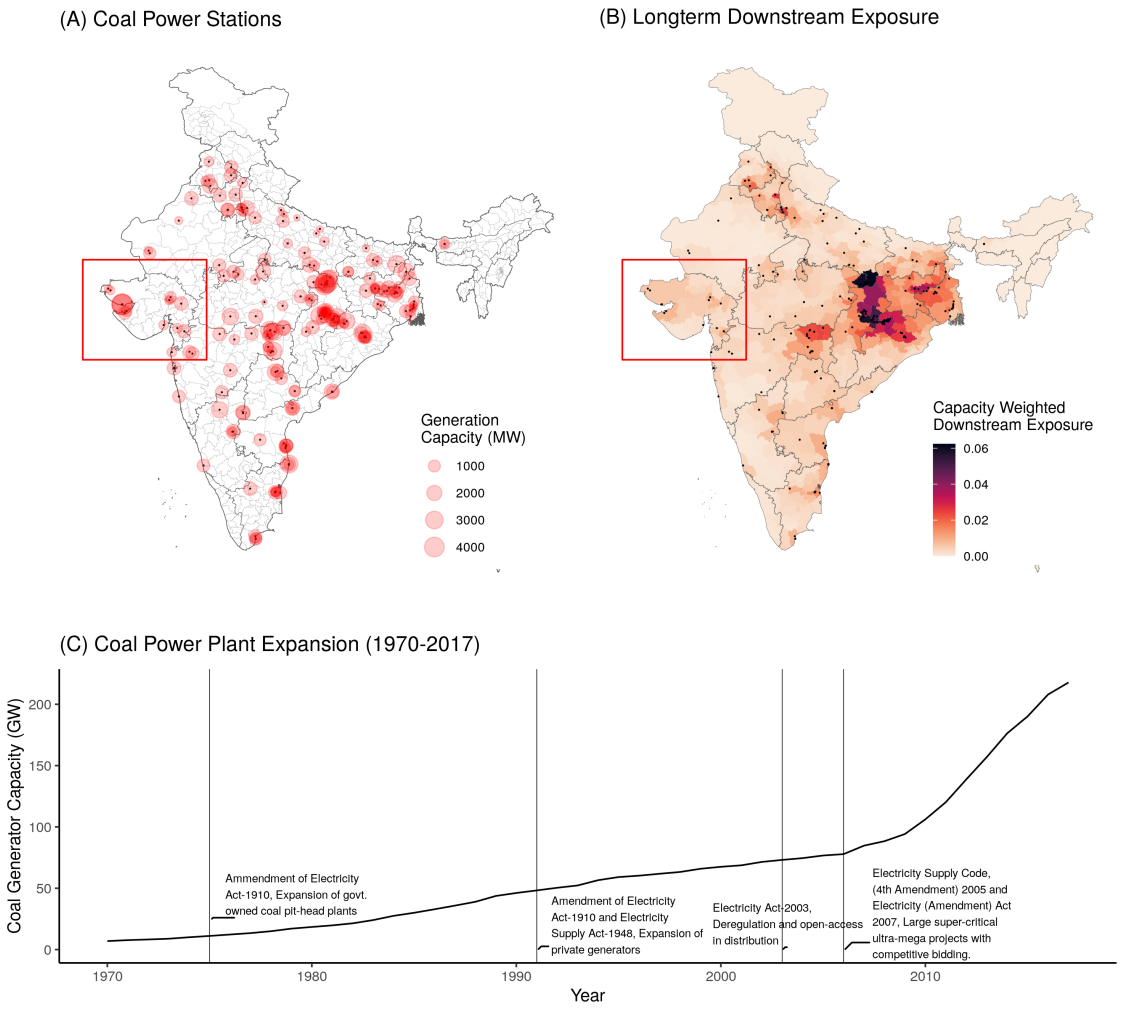


Figure 2: Coal Power Plant Capacity Expansion

(A) Location and capacities of coal fired power plants in India (2017). (B) Long-term cumulative downstream exposure estimated by the HYSPLIT model at the district-level (1970-2017). (C) Time trend of the total installed capacity of coal fired power plants in India (1970-2017). The major policy changes influencing the different policy driven capacity expansion phases are shown along the timeline.

gigawatts (GW), and then summing over the full set of coal stations:

$$GridExposure_{it} = \sum_{j} \left(\left(\frac{EmissionCount_{ijt}}{\sum_i EmissionCount_{ijt}} \right) \times Capacity_{jt} \right) \quad (1)$$

$GridExposure_{it}$ can be interpreted in GW equivalents, since the capacity of plant j is distributed across the grid according to the exposure of each cell to emissions from the plant. Intuitively, if plant capacity is 1 GW and the exposure of a far-away cell is 0.01, then it is as if there is 10 MW of installed capacity in the cell even though, in reality, there is no power plant there.

Third, we aggregate downstream exposure at the district-level. $Exposure_{dt}$ for district d in time period t measures total exposure per cell in a district, where each spatial intersection consists of $GridCount_d$ number of grid cells pro-rated on cell area for partial intersections:

$$Exposure_{dt} = \frac{1}{GridCount_d} \sum_i GridExposure_{it}, \ \forall \ i \text{ spatially intersecting with district } d \qquad (2)$$

To elicit the long-term impacts of power plant exposure, we accumulate $Exposure_{dt}$. The longest cumulation window of fixed length is thirty years since mortality data is available for 2000, 2007, and 2017, and the oldest power plant in our sample was constructed in 1970. Although the start year of cumulation differs in each cross-section (i.e. 1970 for 2000, 1977 for 2007, etc), long-term exposure always refers to three decades of continuous exposure to power plants, which allows for analogous coefficient interpretation in our regressions. Figure 2B shows the spatial distribution of cumulative $Exposure_{dt}$ for the year 2017.

Advantages of a Dispersion Model Generated Pollution Exposure Measure. Our power plant exposure measure offers several advantages that enable credible identification of long-term impacts. First, the exposure measure is entirely model-generated and requires no assumptions about the shape and size of dispersion plumes. In contrast, conventional wind-direction measures assume a plume shape (typically a 45° cone around the wind vector) and length and do not account for diffusion over space.

Second, our exposure measure is well suited for long-term analyses. Since it is based on tracer particle counts, aggregation to any level retains all the spatial distribution and concentration profile information. This allows for credible comparisons of health outcomes in long-run exposed versus non-exposed districts. Conventional measures typically capture within-year seasonal variation in wind direction, which loses substantial explanatory power when aggregated over multiple years (Anderson, 2019; Deryugina et al., 2019).

Third, since HYSPLIT independently computes wind-field driven air transport from each source station, we are able to apportion differential exposure at a pollution species-level (i.e. SO_2, NO_2, $PM_{2.5}$) in the receptor districts attributable to specific source power plants.

Fourth, using our method, spatial contours of the downstream exposure plume can be explicitly demarcated, either through machine learning algorithms or percentile-based cutoffs. We can then correct for potential spatial correlation among districts under the same plume by clustering standard errors at the plume level (see section 3.2). This enables more precise estimation of stan-

dard errors compared to existing literature.

Finally, our approach allows for meaningful counterfactual policy simulations. Since our exposure measure is aggregated from plant-level dispersion runs, we can easily take plants offline, simulate abatement scenarios through capacity adjustments, or simulate different time-trends for coal phase-downs.

3 Research Design

Our dispersion model generates the complete spatial extent and intensity of wind-driven exposure to coal power plant emissions across all districts. However, power plants themselves are not randomly sited, complicating inference from comparing outcomes in exposed versus non-exposed districts. This section describes our research design for credibly estimating long-term impacts of power plant exposure. We first build intuition, then demonstrate the improved predictive accuracy of our exposure measure compared to conventional approaches, and then finally state our formal estimating equations.

3.1 Motivating Facts

Power plants are sited primarily based on geographic suitability. Plant location is endogenous and reflects many factors, especially geographic suitability. This natural resource basis for power generation has been an institutional feature of India's power sector since the 1961-66 Five-Year Plan:

> "Steam power stations should be sited near collieries [coal mines], washeries [water bodies] and oil refineries. All power stations should be inter-connected to form state, zonal or super-grids, so that the energy is pooled and used to the best advantage of the region."(Planning Commission, 1961)

This set the stage for India's "coal-by-wire policy," which sought to site new power plants near the coal deposits. In many countries that underwent rapid expansions of their fleet of coal-fired power plants, including China and the United States, power plants were sited near the population and industrial centers that consumed the largest shares of the generated electricity. This system required extensive reliance on rail networks to ship coal from mines to power plants, a costly endeavor that often more than doubles the delivered cost of coal relative to its cost at the mine mouth. Under the coal-by-wire policy, India pursued power sector expansion by siting plants near the coal mines and deposits, providing the plants' fuel rather than demand centers. Instead of relying on costly transportation of coal over long distances to plants, it relied on transportation of the generated electricity from plants to demand centers by way of the high-voltage transmission networks.

Table 1: Power Plant Placement

	Shift-Share [2008-2017]		Cross Section		
	Annual	Monthly	2017		
	(1)	(2)	(3)	(4)	(5)
	Capacity (MW)	Capacity (MW)	Exposure (no mask)	Exposure (no plants)	Exposure (50km mask)
Proximity to Coal	2.650***	2.640***	20.059**	7.711*	8.214**
	(0.723)	(0.681)	(7.300)	(4.273)	(3.081)
Nearest Coal Area	-0.002	-0.002	-0.047	-0.065	-0.046
	(0.003)	(0.002)	(0.055)	(0.042)	(0.032)
Water Area w/n 25km	0.049**	0.062***	0.774***	0.331	0.245
	(0.020)	(0.021)	(0.161)	(0.256)	(0.157)
Proximity to Coast	-0.022	0.003	-0.853	-1.078	-1.112
	(0.119)	(0.113)	(0.539)	(0.706)	(0.741)
Elevation	-0.020	-0.018	0.081	0.048	0.063
	(0.033)	(0.030)	(0.086)	(0.052)	(0.045)
Slope	-0.176	-0.150	-0.549	-1.303*	-1.054
	(0.369)	(0.366)	(1.658)	(0.667)	(0.660)
District FEs	✓	✓			
State FEs			✓	✓	✓
State × Year FEs	✓	✓			
Month FEs		✓			
Clustering	District	District	Plume	Plume	Plume
N	6290	75208	632	525	605
R^2	0.899	0.894	0.452	0.570	0.622

* $p < .1$, ** $p < .05$, *** $p < .01$. Proximity to coal and coast are measured as inverse distance from the district centroid. Coal and water area are measured as means over district cells. Coefficients on elevation and coal area are multiplied by 1000. Slope and water area are multiplied by 1000 and 10^8, respectively. In columns 1-2, row variables are interacted with predicted state capacity. Columns 3-5 represent cross-sectional regressions in 2017 where rows are uninteracted. Exposure is cumulated over thirty years. Masks define the radius around power plants within which cells are deleted prior to aggregation. Column 4 applies no mask and drops all districts with power plants.

Despite this reform, capacity investment remained stagnant until the mid-2000s, after which the electricity sector unbundled and opened up to private players. Installed coal-fired capacity nearly tripled thereafter between 2008-2017 (see figure 2C).

Table 1 shows the role of the coal-by-wire policy during this dramatic expansion. Columns 1 and 2 show six geographic determinants of yearly and monthly capacity expansions. Each geography variable is fixed and thus enters interacted with installed capacity at the state level. The latter is predicted from pre-period state shares of national capacity to ensure exogeneity[5]. Each coefficient therefore describes the share of state capacity allocated to districts closer to coal deposits,

[5]This type of specification mirrors the standard shift-share design (Goldsmith-Pinkham et al., 2020).

water bodies, and so on.

We find clear evidence of strategic placement. Districts closer to coal deposits, and with more proximate water sources, receive significantly larger shares of state capacity. Conditional on distance to the nearest deposit, coal deposit area does not influence investment. Districts with higher elevation and steeper terrain attract less capacity investment, although estimates are noisy.

Having established that geographically suitable districts attract capacity investment, we can expect that these same districts will be more exposed to power plant pollution either by having their own plant or by being close to one. This introduces an important source of endogeneity near power plants whereby exposure is partially determined by the coal-by-wire policy. In contrast, our exposure measure is arguably exogenous at greater distances since it will be driven by wind field characteristics and not any placement effect *per se*.

Columns 3-5 successively illustrate this endogeneity and our solution to it. Each specification is a cross-sectional regression of cumulative power plant exposure (equation 2) in 2017 on the same six geography variables with state fixed effects. Column 3 shows a strong correlation between exposure and proximity to coal and water—the same features predicting plant placement in columns 1 and 2. Column 4 disentangles the placement effect by dropping power plant districts altogether, leaving coefficients to be estimated mainly off of wind-driven exposure. Correlations between geography and exposure almost entirely vanish. While this approach largely solves our endogeneity problem, it requires dropping 107 districts. Furthermore, the placement effect may still partially operate since some plants are sited at district borders (see figure 3A).

Column 5 presents estimates under an alternative approach that addresses both sample size and border issues. We apply a 50km mask around each power plant, which deletes cell values $GridExposure_{it}$ (equation 1) within a 50km radius. This accounts for spillover coal-by-wire effects and drops only the handful of districts smaller than the mask. Coefficients and precision are nearly equivalent to column 4 and only 27 districts are dropped. We apply this mask throughout our main analysis and vary the radius in the robustness checks.

Districts face unequal exposure, *conditional* on geography. Our research design is careful to account for the natural resource basis for power generation. We control for a range of geographic determinants of plant placement, including distance to the nearest plant itself. Residual variation in long-term exposure between districts thus stems from conditionally exogenous differences in wind-fields and other meteorological characteristics used by HYSPLIT.

The intuition behind our identifying variation is well captured by the experience of two districts on the western coast of Gujarat. Rajkot and Junagadh are the 4th and 7th most populous districts in the state, respectively. They are also neighbours. Neither have their own power plant but both are approximately 150km from the nearest plant cluster (the Salya and Sikka Power Stations) located in nearby Jamnagar district. All three districts are thus expected to share similar ge-

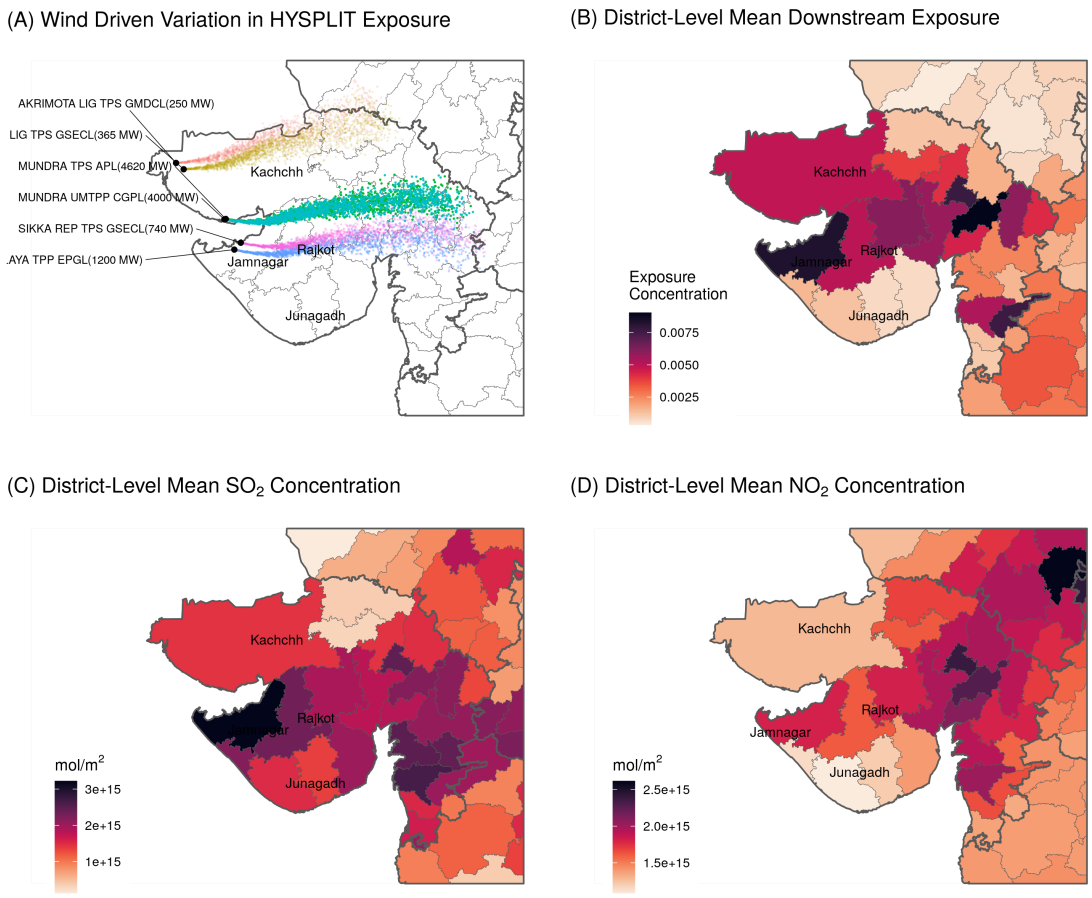


Figure 3: Wind-field driven downstream emission exposure and pollution.

(A) An illustration of the downstream emission dispersion 'plume' estimated by the HYSPLIT model from six power plants located on the western coast of Gujarat for April 2016 (inset from Fig.2). (B) District-level annual mean downstream exposure concentration from HYSPLIT model for 2016. (C) District-level annual mean SO_2 pollution concentration for 2016. (D) District-level annual mean NO_2 pollution concentration for 2016.

ographic characteristics. Figure 3A highlights these districts and shows particle trajectories from each plant in April 2016. Despite their similar development, proximity, and geography, Rajkot and Junagadh face drastically different exposure to the nearby power plants. Rajkot is directly in the exposure path of the Jamnagar plant cluster as well as that of the Mundra Ultra Mega Power Plant in further-away Kutch district (green). Junagadh is completely non-exposed. These differences in exposure arise even when aggregating at the annual level (Figure 3B). It is this variation that we exploit to estimate the causal impact of power plant exposure.

Table 2: Comparison of Predictive Accuracy: $\log(SO_2)$

	Within R-sq.	Relevance Tests	
		Kleibergen and Paap (2006)	Montiel Olea and Pflueger (2013)
	(1)	(2)	(3)
HYSPLIT			
Capacity-weighted Exposure (1990-)	0.350	50.955	49.214
Wind Cone			
Capacity-weighted Cone (1990-)	0.231	16.253	15.242
Periods overlapped by cone (1990-)	0.208	11.672	12.345

Note: each row represents a regression of $\log(SO_2)$ on exposure, controls for plant placement, and state-year fixed effects. Controls include: district area, distance to nearest river, water area, distance to coal deposit, coal area, distance to nearest power plant, elevation, slope, temperature, and nightlight intensity. The first row is clustered at the plume level and the remainder at the state level.

Districts face unequal pollution, *conditional* on geography. HYSPLIT maps the trajectory of tracer particles as opposed to actual air pollutants. But since tracer particles largely follow wind-fields, we expect historically exposed downwind districts to also experience worse air quality today compared to upstream districts that avoided exposure.

The intuition can again be illustrated by the experiences of Rajkot and Junagadh districts. Recall that they are neighbors and equi-distant from the nearest power plant, but only Rajkot is in the downwind exposure path. Figure 3C and D shows gridded annual SO_2 and NO_2 concentrations, respectively, in 2016. The worst *observed* air quality is in the *model-generated* downwind exposure path, which includes Rajkot. In contrast, upwind districts like Junagadh, which are not exposed to power plants, experience relatively lower pollution levels.

This visual evidence implies 1) that our model-driven exposure measure is a strong predictor of actual observed pollution and; 2) that pollution constitutes a key mechanism through which differential exposure translates into health disparities. We test the first claim next, and the second in section 4.

Dispersion model-based exposure predicts pollution better than conventional measures. We demonstrate the advantage of our exposure measure by comparing how accurately it predicts pollution compared to the standard wind direction instrument from the literature. To replicate the standard measure, we obtain monthly gridded wind direction from 1990-2017 at $0.1 \times 0.1^{\circ}$ resolution from the ERA5 satellite product.[6] First, we extract the wind direction at the pixel of each power plant. Second, we construct a quarter-circle with radius 100km centered on the vector. This

[6]We use the ERA5-Land monthly averaged data obtained from the Copernicus Climate Change Service (C3S)

cone defines the exposure area of the power plant. Third, we distribute plant capacity over districts spanned by the cone, weighted by overlap fraction, and then sum over plants. This produces a measure of district exposure to power plants loosely resembling our HYSPLIT-generated measure, but the lack of a dispersion model necessitates strict assumptions about plume shape and diffusion characteristics. Lastly, we accumulate the capacity-weighted cones from 1990 to produce a long-term exposure measure for comparison. We also compute total months in which a district touches an exposure cone as an alternative measure.

Table 2 provides formal diagnostics of predictive performance. The HYSPLIT measure is adjusted to a 1990 start date for consistency. Each row represents a regression of log SO_2 on exposure, controls for plant placement described in table 1, and state-year fixed effects. Columns describe performance metrics. Our main diagnostics derive from two variants of the first stage F-statistic. The Kleibergen and Paap Wald statistic is a heteroskedasticity-robust analog of the first stage F-statistic (Kleibergen and Paap, 2006). Olea and Pflueger (2013) adjust this test to accommodate clustered panel data, which suits our setting. Their simulations imply a critical value of 23 compared to the standard rule-of-thumb level of 10 suggested by Staiger and Stock (1994).

Column 1 shows the within-R^2 from each regression. Our model-based measure explains over 50% more pollution variation compared to the cone-based measure. A more relevant test of performance is whether the parameter mapping exposure to pollution is weakly identified. Column 2 shows heteroskedastic-robust F-statistics. Values range from 12-16 for cone-based measures whereas our measure is three times larger. Although all exceed 10, it should be noted that this rule-of-thumb assumes a homoskedastic setting (Staiger and Stock, 1994). In column 3, where the test and critical value were designed for the same setting, our measure is strongly relevant for predicting pollution and exceeds the critical value (=23) by a wide margin. By contrast, both cone-based measures under-perform and are deemed weak predictors by the test.

The equivalent diagnostics for NO2 pollution are provided in Appendix table A1. The F-statistics for our HYSPLT-based exposure measure is over 6 times that of the cone-based measure.

3.2 Empirical Strategy

Main Specification. Our empirical strategy generalizes the experiences of Rajkot and Junagadh discussed in the previous section. We compare health outcomes across districts with different historic exposures to power plants, controlling for state fixed effects and various plant placement determinants. We do this for three separate periods several years apart. Thus, our estimation strategy exploits cross-sectional variation, giving rise to plausible control groups for historically exposed districts: other districts within the same state sharing the same geographic characteristics, located at equal distance to power plants, but historically less exposed to these plants due to plausibly random differences in their wind fields.

We estimate the following specification on three stacked cross-sections:

$$Mortality_{dt} = \beta_1 CmlExposure_{dt} + \beta_2 X_{dt} + \gamma_{st} + \mu_{dt}, \tag{3}$$

where $Mortality_{dt}$ is either neo-natal, infant, or child deaths per 1000 live births in district d during cross-section $t \in \{2000, 2007, 2017\}$. $CmlExposure_{dt}$ denotes the amount of cumulative exposure to all upwind coal power plants in the past three decades (i.e. between $t-30$ and t) in district d. X_{dt} is a vector of fixed covariates determining plant placement described in table 1. The majority of covariates are constant in each t (i.e. slope, distance to nearest coal deposit) except for distance to nearest power plant, which changes as new plants are constructed. State-year fixed effects, γ_{st}, are crucial for ensuring that cross-district comparisons are made within states separately in each period. Without the interaction of year fixed effects, the identifying variation would conflate long- and short-term variation.

The coefficient of interest, β_1, captures the modern-day health impacts of district d being situated downwind from historic power plants and therefore being continuously exposed to worse air quality for three decades. Whereas location-by-time fixed effects typically identify short-term deviations, in our case β_1 represents the average of three separate cross-sectional comparisons.

Threats to Identification. Our main identifying assumption is that historic exposure to power plant emissions is orthogonal to other correlates of current health outcomes. Although we cannot test this assumption directly, we provide several pieces of corroborating evidence. First, the institutional context demonstrates a natural resource basis for plant placement rather than a decision tied to pre-existing health or economic trends. Second, we test for pre-trends directly and show that, in the absence of a power plant, districts show no evidence of a pollution trend. Third, we rule out adaptation and endogenous sorting as alternative mechanisms.

Standard Error Clustering. With each simulated "puff" from the dispersion model, gridded exposure values are spatially correlated across cells through unobserved physical formulas governing gas movement. This correlation persists when aggregating exposure to the district level. The ideal solution is to cluster standard errors at the plume level, which allows residual correlation across districts and over time within plume boundaries.

We implement a novel method that demarcates plumes using unsupervised machine learning. We first run a k-means clustering algorithm on each exposure grid (equation 1) to partition cells into groups. The algorithm identifies subgroups, which we call plumes, in the raster within which exposure values are as similar as possible according to a euclidean distance metric (MacQueen et al., 1967). We then assign a common cluster ID to districts spanned by the same plume. Districts spanned by multiple plumes are assigned the one with the highest overlap fraction.

Table 3: Reduced Form: Long Term Impact of Historic Exposure on Mortality

	Neo-Natal	Infant	Child
	(1)	(2)	(3)
Capacity-weighted Exposure	0.583***	0.883***	1.284***
	(0.191)	(0.301)	(0.387)
Controls	Yes	Yes	Yes
Y Mean	29.306	46.897	59.020
X SD	0.454	0.457	0.454
State × Year FEs	✓	✓	✓
Clustering	Plume	Plume	Plume
N	1784	1792	1786
R^2	0.810	0.828	0.854

* $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the district-annual level for 2000, 2007, and 2017. Mortality is measured per 1000 live births. Exposure is cumulative over 30 years, includes a 50km mask, and standardized. All specifications include controls for: district area, distance to nearest river, water area, distance to coal deposit, coal area, distance to nearest power plant, elevation, slope, temperature, and nightlight intensity.

The advantage of this approach is that all cells and districts are spanned by a plume. The limitation is that the researcher pre-sets the number of plumes; we use 25. We test a different plume-demarcation approach in the robustness checks that relies on percentile based cutoffs.

4 Results

We present four main results about the health impacts long-term exposure to coal power plants. First, districts historically exposed to more power plant pollution over three decades experience higher present day mortality. Second, worse air quality in exposed districts is the key mechanism driving additional deaths. Third, negative health effects of power plant exposure are more pronounced in urban districts. Lastly, exposure in-utero is the most critical period influencing present-day mortality. Section 5 shows a range of robustness tests as well as evidence ruling out two competing explanations for our findings.

4.1 Historically Exposed Districts Experience Higher Mortality

Stacked Cross-Section Results. Table 3 shows the impacts of historic exposure on neo-natal, infant, and child mortality. Neo-natal mortality is defined as death before 28 days old, infant mortality as death before one years old, and child mortality before five years old. The state-by-year fixed effects ensure estimates are based off of three separate cross-sectional comparisons in 2000, 2007, and 2017. $CmlExposure_{dt}$ is standardized for ease of interpretation: a one unit increase in the

independent variable corresponds to a one standard deviation increase in cumulative capacity-weighted exposure to power plants.

Historic exposure to coal power plants significantly increases present-day mortality. Column 1 shows that districts cumulatively exposed to an additional standard deviation of power plant emissions over three decades experience 0.6 more neonatal deaths per 1000 births, representing 2.0% of the mean. The equivalent increase for infant and child mortality is 1.9% (column 2) and 2.2% (column 3), respectively. All three estimates are statistically significant at 1%.

To the best of our knowledge, there are no other estimates of the long-term mortality impacts of coal power plants to which we can compare our results. However, we can provide at least two sanity checks on whether the magnitude of our estimates are reasonable. First, we transform our long-term estimates into analogous short-term estimates, for which there is an ample evidence base for comparison. Barrows et al. (2019) find that a 1 GW increase in yearly wind-driven power plant exposure causes 2.9 additional infant deaths per 1000 live births across Indian districts[7]. Here, the equivalent annual impact based on the estimate in column 2 is 12.8 additional infant deaths per year[8]. The discrepancy may be explained by differences in study period and, more importantly, in the construction of exposure.

Second, we calculate how historic power plant exposure contributes to India's mortality trend. Average district-level neonatal mortality declined from 36 to 22 deaths per 1000 live births between 2000-17. At the same time, the average district faced a 0.3 GW (or, 0.7 standard deviation) historical exposure to power plants. Table 3 column 1 implies that this led to 0.4 ($= 0.7 \times 0.583$) additional deaths, suggesting that the mortality decline in the average Indian district suffered a 3 percent ($= 0.4/14$) setback due to historic power plant exposure.

4.2 Pollution is the Key Mechanism

Pollution pre-trends. Since wind characteristics drive our dispersion model, it is natural to presume that highly exposed districts also experience worse air quality due to their downwind orientation towards power plants. Visual evidence from districts differentially exposed to power plants in Western Gujarat corroborates this (Figure 3C,D).

We establish that pollution is a key mechanism linking power plant exposure to health by first testing for pre-trends. Figure A2, which we used previously to demonstrate pollution reduction during plant end-of-life, also shows no evidence of pre-trends. The date of commissioning is

[7] Barrows et al. (2019) report that 1 standard deviation of exposure increases infant mortality by 0.8. The standard deviation of downwind exposure reported in their summary statistics is 0.274, implying that 1 GW causes 0.8/0.274=2.9 infant deaths per 1000 live births.

[8] Column 2 implies that 1 SD of exposure (=0.457 GW) causes 0.883 additional present-day infant deaths. Thus, 1 GW of exposure causes 0.883/0.457=1.93 additional deaths. This implies 1.93/30=0.064 deaths per thousand live births for every 1GW/year. Given 200GW of installed capacity across India, this implies roughly 200*0.064=12.8 additional infant deaths per 1000 live births per year.

Table 4: Impact of Historic Exposure on Log Pollution

	(1) SO2	(2) NO2	(3) PM2.5
Capacity-weighted Exposure	0.025***	0.010***	0.480
	(0.003)	(0.001)	(0.338)
Controls	Yes	Yes	Yes
Y Mean	0.052	0.063	43.528
X SD	0.461	0.461	0.461
State × Year FEs	✓	✓	✓
Clustering	Plume	Plume	Plume
N	1830	1830	1830
R^2	0.734	0.863	0.885

* $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the district-annual level for 2000, 2007, and 2017. SO_2 and NO_2 are measured in Dobson Units (1 DU= 2.69×10^{16} molecules/cm2). $PM_{2.5}$ is measured in $\mu g/m^3$. Exposure is cumulative over thirty years, includes a 50km mask, and standardized. All specifications include controls for: district area, distance to nearest river, distance to coal deposit, distance to nearest power plant, elevation, slope, temperature, and nightlight intensity.

the omitted group. Coefficient estimates 1-5 years prior to plant commissioning are statistically indistinguishable from zero, implying that the timing of plant opening is exogenous to unobserved determinants of pollution. If plant placement was correlated with local demand, for example, then pollution would trend in the direction of district GDP even during the pre-period.

Historic Exposure and Ambient Air Pollution. Table 4 formally tests the pollution mechanism by estimating equation 3 with pollution as the outcome. There is a large and precise effect of historic power plant exposure on district-level SO_2 and NO_2 concentrations. Column 1 shows that a one standard deviation increase in historic exposure increases present-day SO_2 concentrations by 50 percent of the mean. The equivalent increase in present-day NO_2 is 16 percent (column 2).

Column 3 shows that $PM_{2.5}$ concentrations are surprisingly no different between historically exposed and less-exposed districts. One explanation for the discrepancy is that NO_2 and SO_2 are known to be associated with fossil fuel combustion whereas coal-fired power plants are not a major $PM_{2.5}$ source in India. Recent source apportionment studies show that coal-fired power plants account for 30 percent and 50 percent of India's NO_2 and SO_2 emissions, respectively (Lu and Streets, 2012; Lu et al., 2013), but just seven percent of $PM_{2.5}$ emissions (Venkataraman et al., 2018). Cross-district variation in $PM_{2.5}$ is likely too noisy to detect the impact of power plants.

4.3 Health Effects are More Pronounced in Urban Areas

Our main results represent an average across Indian districts. An important question is whether urban form moderates the relationship between historic power plant exposure and early-life mor-

Table 5: Heterogeneity

	Mortality			Pollution		
	(1) Neonatal	(2) Infant	(3) Child	(4) SO2	(5) NO2	(6) PM2.5
Capacity-weighted Exposure	0.512**	0.762**	1.131***	0.270***	0.115***	0.003
	(0.196)	(0.306)	(0.393)	(0.033)	(0.008)	(0.007)
Capacity-weighted Exposure × Urban (=1)	0.641	1.320*	1.868**	0.075	-0.006	-0.015
	(0.436)	(0.701)	(0.851)	(0.062)	(0.022)	(0.020)
Urban (=1)	-0.828**	-1.407**	-1.792**	-0.049	-0.035	-0.052***
	(0.363)	(0.560)	(0.684)	(0.037)	(0.022)	(0.019)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
X SD	0.454	0.457	0.454	0.461	0.461	0.461
State × Year FEs	✓	✓	✓	✓	✓	✓
Clustering	Plume	Plume	Plume	Plume	Plume	Plume
N	1784	1792	1786	1830	1830	1830
R^2	0.809	0.827	0.853	0.678	0.860	0.911

* $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the district-annual level for 2000, 2007, and 2017. Mortality is measured per 1000 live births. SO_2 and NO_2 are measured in Dobson Units (1 DU= 2.69×10^{16} molecules/cm2). $PM_{2.5}$ is measured in $\mu g/m^3$. Exposure is cumulative over thirty years, includes a 50km mask, and standardized. Urban (=1) is a dummy for whether more than half the district population is urban. All specifications include controls for: district area, distance to nearest river, water area, distance to coal deposit, coal area, distance to nearest power plant, elevation, slope, temperature, and nightlight intensity.

tality. This heterogeneity is especially important in India, home to some of the most population dense cities in the world. We investigate this question by documenting whether urban and rural districts experience different mortality outcomes from the same level of exposure to power plants.

We estimate our main specification (equation 3) with an interaction dummy for urban districts, equal to one if more than half the population is urban. Many other confounding differences between urban and rural districts are accounted for by our geography and distance-to-plant controls. The interaction term thus identifies heterogeneity by population exposure holding other agglomeration effects constant.

Table 5 presents our heterogeneity results. Columns 1-3 show that the effect of power plant exposure on early-life mortality is more pronounced in urban districts. The same level of exposure triggers 1.3 and 1.9 more infant and child deaths per 1000 live births, respectively, in urban districts compared to rural ones. To put this in perspective, urban non-exposed districts are healthier than their rural counterparts (third row), perhaps because of higher incomes and better health care delivery in cities. Long-term exposure to power plants almost entirely erases this "urban advantage".

Differences in population exposure is the most likely mechanism driving heterogeneity in columns 1-3. If an urban and rural district are equidistant to a power plant and face the same

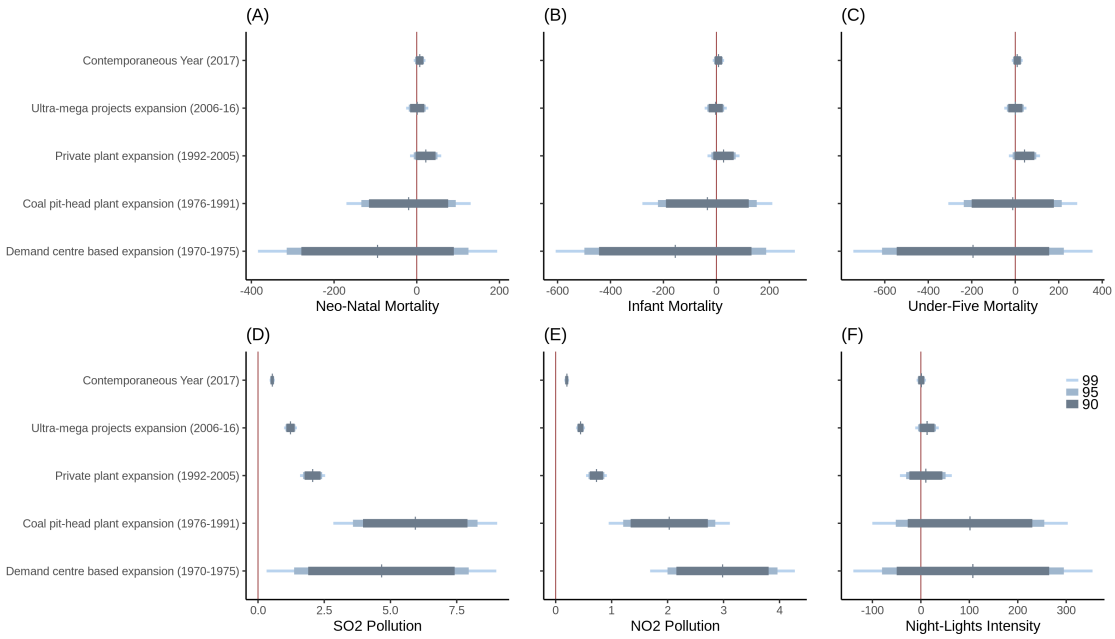


Figure 4: Impact of historic coal expansion phases

Note: y-axis is mean exposure within the time intervals corresponding to salient policy driven coal power expansion phases. Each panel shows coefficient estimates on mean exposure in the respective time intervals from five different regressions. Exposure includes a 50km mask and is standardized. All specifications include controls for: district area, distance to nearest river, distance to coal deposit, distance to nearest power plant, elevation, slope, night lights, rainfall, and temperature.

pollution exposure, the urban district will still experience worse mortality because higher population density implies greater population exposure.

We verify this mechanism by ruling out one remaining contender: differential pollution effects. If higher mortality in urban districts were driven by higher pollution in urban districts, then we should see a positive coefficient on the interaction, with pollution as the outcome. This is unlikely since our dispersion model is powered by wind and other meteorological characteristics, such that power plant pollution is equally likely to land in urban versus rural districts. The null effects of the interaction term in Columns 4-6 confirm this hypothesis. Urban form moderates health effects through differential population exposure, not pollution.

4.4 In-utero Exposure and the Period of Private Power Plant Expansion are the Most Critical

We have shown that districts historically exposed to power plants over five decades experience significantly higher present-day neonatal, infant, and child mortality. A natural question is: at what point during this historical period is exposure most critical for determining current health outcomes? We answer this by unraveling historic cumulative exposure across distinct phases of policy driven growth in coal fired power plants.

We decompose cumulative exposure in equation 3 into multiple time bins to investigate persistence. The bins spans four distinct time periods marked by policy changes influencing the growth of power generators in India. In addition, a bin corresponding to the most recent year (2017) is included, this is also the contemporaneous year for mortality and pollution outcome measures. The contemporaneous period is also the period of in-utero exposure for 2017 mortality data. Cumulative exposure is converted to a per-year average (i.e. total exposure divided by number of years in the bin) for consistent interpretation. All fixed effects and controls are the same as equation 3.

Figure 4 shows the decomposition with multiple outcomes. There are four noteworthy results. First, exposure in-utero is a critical period for influencing infant survival in the first 28 days (neonatal mortality) (Panel A), first month (infant mortality) (Panel B), and under-five years mortality (Panel C).

Second, the decomposition along the power plant expansion phases show that the period of growth dominated by private independent power producers (1992-2005), maintains a significant persistent influence on neo-natal, infant, and under-five mortality in 2017. This is the period following deregulation of power generation for private participation (1991 onward). Exposure during other historic periods shows no persistence on present-day mortality outcomes.

Third, the decomposition shows that the contemporaneous year and all the historic exposure bins influence the present day SO_2 (Panel D) and NO_2 pollution (Panel E) to varying degrees. The historic persistent effect implies that sites of power plant construction in the past continues to influence the present-day plant placements, and through this channel influences the current year cross sectional variation in power plant pollution. Notably, the large effects of the 1975-91 period on current period power-plant pollution indicates that the geographic determinants of coal pit-head siting of plants continues to influence placement of new power capacity within close proximity to the older plants, creating spatially concentrated clusters of coal power generators.

Jointly reading the time period decomposition of mortality effects and pollution effects, we can infer that the present-day mortality outcome is likely influenced the most by power plant clusters that where constructed during the growth phase dominated by privately owned power plants between 1991 and 2005.

Lastly, the decomposition shows no trend when economic activity, measured by nightlight

Table 6: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
Capacity-weighted Exposure	0.618***	1.237***	0.576***	0.583**	0.790
	(0.223)	(0.366)	(0.197)	(0.233)	(1.325)
Dataset	GBD	GBD	GBD	GBD	NFHS
Mask	50km	None	30km	50km	50km
Plant Districts	Yes	No	Yes	Yes	Yes
Y Mean	29.306	29.267	29.161	29.306	25.593
X SD	0.454	0.423	0.454	0.454	0.495
State FEs	✓				✓
State × Time Trend	✓				
State × Year FEs		✓	✓	✓	
Clustering	Plume	Plume	Plume	Pctile	Plume
N	1784	1613	1834	1784	503

* $p < .1$, ** $p < .05$, *** $p < .01$. Data in Columns 1-5 are from the Global Burden of Disease study used throughout the analysis. Column 5 is from the NFHS-IV survey. Mask refers to a radius around each plant within with cell values are omitted. Specifications with plant districts include the district with the power plant in the sample. Exposure is cumulative over thirty years and then standardized. All specifications include controls for: district area, distance to nearest river, distance to coal deposit, distance to nearest power plant, elevation, slope, temperature, and nightlight intensity.

intensity, is the outcome (Panel F). This helps rule out adaptation as an alternative mechanism (we show this more formally in section 5.2). If capacity expansions in the previous year triggered exposed districts to invest in adaptive infrastructure, then we would expect these districts to be more developed than non-exposed districts. The null effect for all the time period bins in Panel F shows that this is not the case. We would also expect no difference in present-day mortality if exposed districts adapted. The positive and significant in-utero effect and private expansion period effect shows that this is also not the case.

5 Ruling Out Alternative Explanations

Our empirical strategy carefully assigns districts to treatment and comparison groups that plausibly differ only in their wind-fields relative to thermal power plants. This section first establishes the credibility of our main estimates through a range of robustness checks and then rules out two alternative mechanisms. We show robustness estimates for neonatal mortality to reduce the number of specifications to report.

5.1 Robustness Checks

Robustness to State-Time Trends. Table 6 column 1 shows results with state fixed effects and a linear state-time trend. Even though time variation is exploited, the exposure coefficient is very simi-

lar to our main specification with state-by-year fixed effects. This is likely because our distance-to-plant control and the state-time trend captures endogenous capacity additions over time, leaving column 1 to be identified off of mainly cross-sectional variation as in our main estimates.

Removing Power Plant Affected Areas. Our 50km mask helps isolate exposure to power plant pollution from the impact of power plant placement per se (see section 3.1). Nevertheless, we show results from two alternative specifications that separate placement and exposure. Rather than applying a mask, column 2 drops power plant districts altogether. Neonatal mortality in these emissions-exposed districts, but without power plants themselves, remains significantly higher than non-exposed districts. The point estimate nearly doubles in magnitude.

Column 3 replicates our main specification but reduces the mask radius to 30km. Estimates are virtually unchanged. The number of observations slightly increases since districts larger than a 30km master but smaller than a 50km mask are added to the sample.

Alternative Clustering. Throughout the analysis, errors are clustered at the plume level, where plumes are demarcated through a k-means algorithm. Although entirely data-driven, this approach requires arbitrarily pre-setting the number of plumes. We test robustness to an alternative clustering approach based on percentile-based cutoffs. A plume is defined as the union of cells in the 95th percentile of exposure values (red cells in Figure A1). A common cluster ID is assigned to districts spanned by the same plume. Districts not covered by a plume retain their district ID as the cluster. This is the main caveat since the unobserved meteorology governing dispersion is obviously insensitive to borders.

Our results are robust to clustering on this alternative plume definition. Standard errors in column 4 are slightly higher than our main estimates but the coefficient retains significance at 5 percent.

Estimates from an Alternative Dataset. The GBD mortality data used in this paper compiles multiple sources, namely the population census, civil registration system, and large household surveys (Dandona et al., 2020). We replicate our results on one of these intermediate datasets independently to verify that our estimates are not an artifact of the secondary data generated by Dandona et al. (2020).

We test our results on the 2015-16 NHFS-IV survey, which covers nearly 600,000 households across 29 states and is representative at the district level. The birth history module asks eligible women about all births in the past, including survival and age at death. We use this to compute district-level neonatal mortality. Table 6 column 5 replicates our main estimating equation using the 2014 birth histories. Since this is the year prior to the survey, birth history responses should suffer from the least recall bias.

Table 7: Adaptation to Long Term Power Plant Exposure

	Adaptation				Sorting	
	(1) Literacy	(2) Poverty	(3) Consump.	(4) Service Lbr	(5) Immigration	(6) Emigration
Capacity-weighted Exposure	0.009	0.014*	-162.579	-0.057	-0.215	1.051
	(0.007)	(0.007)	(181.490)	(0.052)	(0.233)	(0.654)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Data Source	Census	SECC	SECC	EC	Census D3	IHDS
Year	2011	2012	2012	2013	2011	2012
Y Mean	0.618	0.331	17838.204	0.759	5.069	23.088
State FEs	✓	✓	✓	✓	✓	✓
Clustering	Plume	Plume	Plume	Plume	Plume	Plume
N	599	599	599	598	605	330
R^2	0.588	0.707	0.727	0.692	0.497	0.519

* $p < .1$, ** $p < .05$, *** $p < .01$. Each column represents a cross-sectional regression from a variety of data sources and years. Columns 1-6 are cross-sectional regressions at the district level. Column 1 is the district literacy share. Column 2 is the mean village poverty rate in a district. Column 3 is mean per capita consumption across urban towns and rural villages. Columns 4 and 5 are log of total employment in the manufacturing and service sector plus 1, respectively. Column 6 is log of total light intensity across all pixels in a district. Column 7 is the district population share consisting of in-migrants who moved with their whole family. Exposure is cumulative over thirty years, includes a 50km mask, and standardized. All specifications include controls for: All specifications include controls for: temperature, rainfall, nightlights, and distance to nearest power plant.

We continue to find a positive neonatal mortality effect using NFHS-IV, although precision is lost. Note, however, that the same parameter in the 2017 cross section of our main analysis (column 7 table A2), the closest period for comparison, is also imprecisely estimated. We posited that pollution reduction from offlining vintage plants can explain the imprecision.

5.2 Adaptation and Endogenous Sorting

Differential Development. Our reduced form estimates of the impact of historic exposure to power plants are net of any adaptation that may have occurred along the way. We partially ruled out this mechanism in Figure 4D, which shows that changes in power plant exposure in various historical periods generated no persistent changes in economic activity from that point on.

We corroborate this evidence further with four additional measures of economic development from multiple data sources. We replicate equation 3 for the year in which each development indicator is provided. All measures are provided through the SHRUG database, and we list underlying data sources in the Table 7 footer.

We find little evidence of differential development. Table 7 columns 1-4 show that districts historically exposed to power plants *do not* have higher present-day literacy rates, higher consumption per-capita, nor greater service sector employment than less exposed districts. Exposed

districts are slightly poorer in the long-run (column 2), however the effect is small in magnitude and weakly significant. Taken together, this evidence bolsters our claim that the long-term effects of power plant exposure on health operate through pollution and not differential development.

Selective Migration. Endogenous sorting presents another competing explanation for our findings. If individuals sort on health or wealth in response to power plant exposure, then our mortality estimates reflect both a pollution and a migration effect. Heblich et al. (2016) show path-dependent sorting following power plant exposure in early 19th century England. We expect little bias from sorting in India, however, due to high socio-economic barriers to migration and historically low migration rates (Munshi and Rosenzweig, 2016).

The 2011 Census D3 tables completely characterize the migrant population in each district of India. We compute the immigrant share of the district population having arrived with their whole family and use it as an outcome in equation 3. If historic exposure drives families to move into cleaner districts, then we expect $\beta_1 < 0$ (i.e. less migration into dirty districts compared to clean ones). Table 7 column 5 shows no evidence of selective migration. Historic power plant exposure does not generate any statistically distinguishable differences in present-day migrant shares compared to less-exposed districts.

Since the Census D3 tables do not disclose migrant origin, we measure whether exposed districts experience greater out-migration using the IHDS 2012 survey. The survey covers 42,000 households across all states. We define the district migration rate as the proportion of households in a district having at least one migrant that has been away for over six months. A limitation is that we cannot observe permanent out-migration of the whole family. The result in Table 7 column 6 provides additional evidence against the endogenous sorting mechanism. Historically exposed districts do not feature relatively higher present-day out-migration rates.

Overall, these results imply that historic exposure does not trigger path-dependent sorting. This rules out endogenous migration as an alternative explanation and once again supports our claim that pollution is the key mechanism driving health disparities between exposed and non-exposed districts.

6 Conclusion

Coal, which is being phased out of the energy system across the developed world, remains a dominant fuel for electricity generation in many developing countries. In India, coal-fired power plants generated 63% of total electricity in 2016 (Shearer et al., 2017). Neighboring China similarly depends on coal for 60% of its electricity. Although electricity access can deliver salient economic benefits, this paper focuses on environmental costs. Until now, previous studies have documented short-term negative health effects from exposure to coal power plants using monthly or annual

panel data. In contrast, we provide novel, long-term evidence on the impact of historic exposure to coal-fired power plants in India over three decades.

We find that districts historically exposed to more power plant emissions, accumulating over thirty years, experience higher present-day neonatal, infant, and child mortality. To arrive at this result, we developed a novel method for measuring historic emissions exposure using a dispersion model that maps wind-driven particle trajectories from every power plant in India. This allows measurement of power plant exposure, and health effects, manifesting in districts without power plants, and potentially hundreds of kilometers away from one. Importantly, we verify that our mortality estimates are driven by higher local pollution in historically exposed districts, and not through differential adaptation or sorting. Lastly, we unravel historic exposure and find that present-day mortality is most sensitive to plant clusters formed during 1992-2005, the period during which the electricity market opened to private generation companies.

In terms of magnitudes, our neonatal estimate implies an average of 15 additional neonatal lives lost in districts facing higher historic exposure to power plants[9]. Since all districts are potentially exposed, according to our definition, this amounts to 15×640=9,600 additional deaths on account of historic power plant exposure. In terms of costs, this translates to approximately $USD 3 billion using the value of a statistical life estimate designed specifically for India by Gulati et al. (2021)[10].

[9]The coefficient in 3 column 1 implies 1 SD (=0.457 GW) of historic exposure increases neonatal mortality by 0.583 deaths per 1000 live births. In our data, average historic exposure is 0.323 GW, or 0.323/0.457=0.71 SD, and average number of live births is 36,524. Thus, the average lives lost per district is 0.71*0.583*36.524=15.1

[10]Gulati et al. (2021) estimate a VSL of $USD 312,663 for India in 2013. Details in the Materials and Methods section of their paper.

A Appendix

A.1 Appendix Figures

Figure A1: Plume Clusters

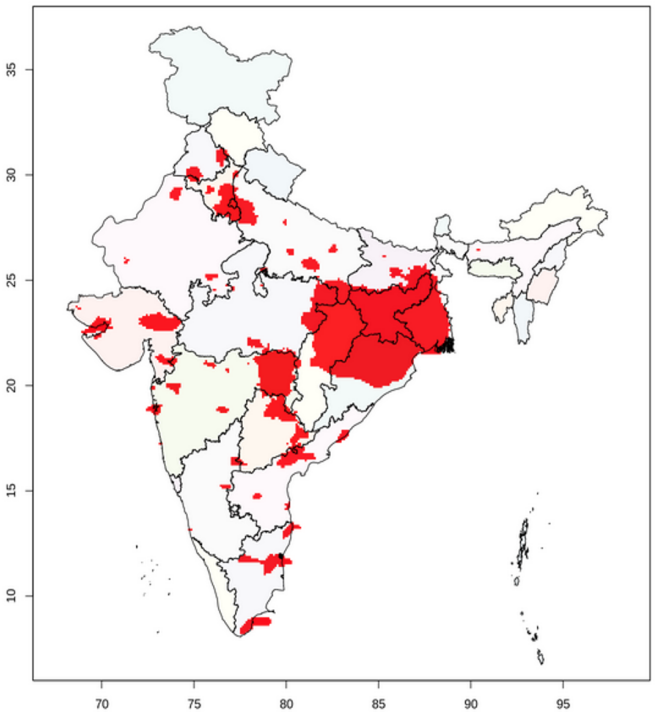
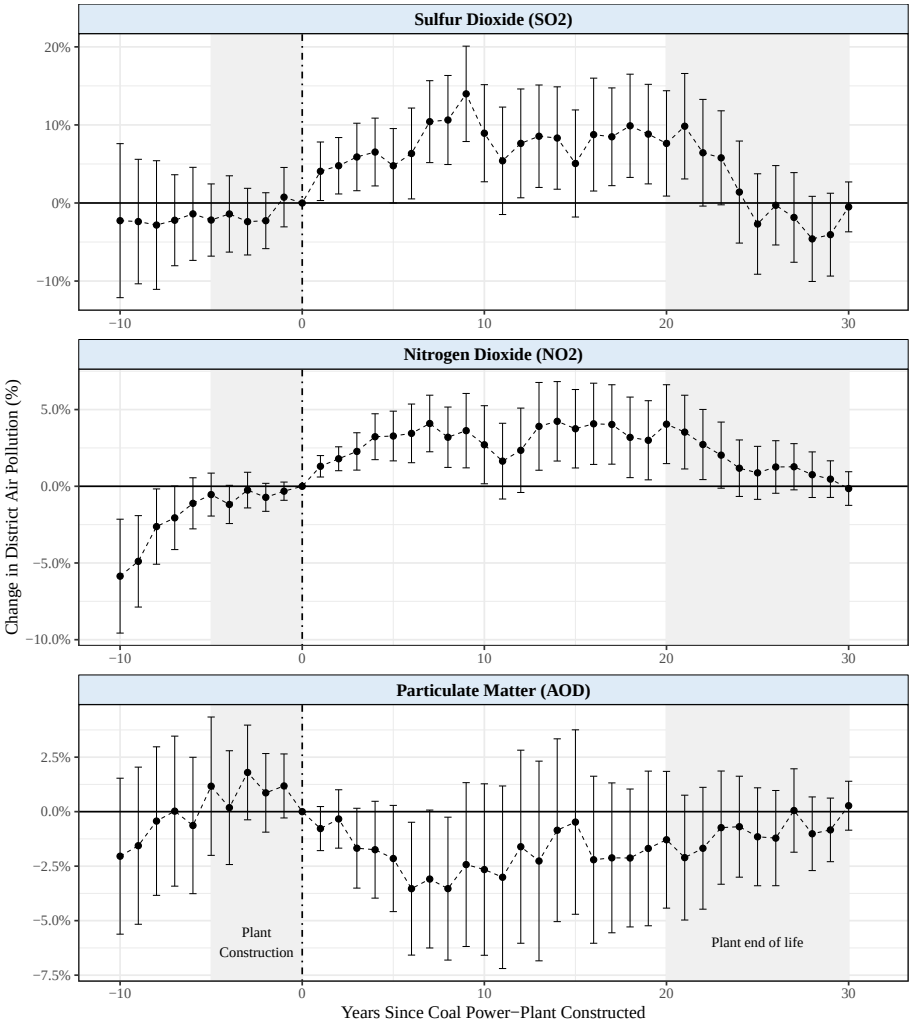


Figure A2: Coal power-plant impact on district air-pollution



We construct a data panel where each power-plant opening is treated as a new event. Districts with multiple plant opening events are treated independently by creating district-event-year ids for each district and stacked to create a panel. With this panel, the time-impact of coal power-plant on district air-pollution (SO_2, NO_2, and Particulate matter) is estimated using a two-way fixed-effects regression specification of the form:
$Log(Y_{it}) = \alpha_i + \gamma_t + \Sigma_{\tau=-10}^{30} \beta_\tau D(\tau) + x_{it} + \epsilon_{it}$. Where Y_{it} are the district pollution outcomes, and the coefficients of interest plotted here are β_τ on the elapsed event-time dummies $D(\tau)$ for $\tau = -10$ *to* 30 years. We also include the existing installed power-plant capacity x_{it} in the specification as control. Error-bars represent 95% confidence.

A.2 Wind-field Emission Dispersion Instrument

Table A1: Comparison of Predictive Accuracy: log(NO_2)

	Within R-sq.	Relevance Tests	
		Kleibergen and Paap (2006)	Montiel Olea and Pflueger (2013)
	(1)	(2)	(3)
HYSPLIT			
Capacity-weighted Exposure (1990-)	0.598	145.209	146.822
Wind Cone			
Capacity-weighted Cone (1990-)	0.542	22.202	22.791
Periods overlapped by cone (1990-)	0.545	43.036	41.748

Note: each row represents a regression of log(NO_2) on exposure, controls for plant placement, and state-year fixed effects. Controls include: district area, distance to nearest river, water area, distance to coal deposit, coal area, distance to nearest power plant, elevation, slope, temperature, and nightlight intensity. The first row is clustered at the plume level and the remainder at the state level.

A.3 Cross-Sections

Table A2: Cross-sectional Long Term Impact of Historic Exposure on Mortality

	2000			2007			2017		
	(1) Neo-natal	(2) Infant	(3) Child	(4) Neo-natal	(5) Infant	(6) Child	(7) Neo-natal	(8) Infant	(9) Child
Capacity-weighted Exposure	0.772 (0.721)	1.315 (1.185)	2.681 (1.600)	1.194*** (0.356)	1.842*** (0.547)	2.589*** (0.694)	0.313 (0.205)	0.423 (0.317)	0.572 (0.379)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Y Mean	36.681	60.658	79.895	28.711	45.569	56.249	22.422	34.316	40.696
X SD	0.229	0.229	0.229	0.350	0.351	0.350	0.622	0.626	0.621
State FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	598	600	598	597	599	597	589	593	591
Clustering	Plume	Plume	Plume	Plume	Plume	Plume	Plume	Plume	Plume
R^2	0.762	0.768	0.795	0.745	0.756	0.772	0.728	0.745	0.748

* $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the district-annual level for 2000, 2007, and 2017. Mortality measured per 1000 live births. Exposure is cumulated over thirty years, includes a 50km mask, and standardized. All specifications include controls for: district area, distance to nearest river, water area, distance to coal deposit, coal area, distance to nearest power plant, elevation, slope, ruggedness, temperature, and nightlight intensity.

Table A3: Cross-sectional Impact of Historic Exposure on Pollution

	2000			2007			2017		
	(1) SO2	(2) NO2	(3) PM2.5	(4) SO2	(5) NO2	(6) PM2.5	(7) SO2	(8) NO2	(9) PM2.5
Capacity-weighted Exposure	0.023*** (0.006)	0.011*** (0.002)	-0.295 (1.107)	0.022*** (0.002)	0.009*** (0.001)	0.183 (0.825)	0.026*** (0.004)	0.010*** (0.001)	0.588 (0.386)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Y Mean	0.040	0.056	35.155	0.045	0.059	42.320	0.070	0.073	53.235
X SD	0.230	0.230	0.230	0.354	0.354	0.354	0.632	0.632	0.632
State FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	613	613	613	612	612	612	605	605	605
Clustering	Plume	Plume	Plume	Plume	Plume	Plume	Plume	Plume	Plume
R^2	0.602	0.858	0.853	0.659	0.850	0.854	0.752	0.853	0.900

* $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the district-annual level for 2000, 2007, and 2017. SO_2 and NO_2 are measured in Dobson Units (1 DU$= 2.69 \times 10^{16}$ molecules/cm2). $PM_{2.5}$ is measured in $\mu g/m^3$. Exposure is cumulative over thirt years, includes a 50km mask, and standardized. All specifications include controls for: district area, distance to nearest river, distance to coal deposit, distance to nearest power plant, elevation, slope, ruggedness, temperature, and nightlight intensity.

References

Anderson, Michael L, "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality," *Journal of the European Economic Association*, 10 2019, *18* (4), 1886–1927.

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.

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad, "Development Research at High Geographic Resolution," 2021.

Authority, Central Electricity, "Report on Minimisation of Water Requirement in Coal Based Thermal Power Stations," Technical Report, Central Electricity Authority, New Delhi January 2012.

Barrows, Geoffrey, Teevrat Garg, and Akshaya Jha, "The Health Costs of Coal-Fired Power Plants in India," *IZA DP No. 12838*, 2019.

Blakeslee, David, Ram Fishman, and Veena Srinivasan, "Way down in the hole: Adaptation to long-term water loss in rural India," *American Economic Review*, 2020, *110* (1), 200–224.

Central Electricity Authority, "Review of Land Requirement for Thermal Power Stations," Technical Report 2010.

Chay, Kenneth Y and Michael Greenstone, "The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession," *The quarterly journal of economics*, 2003, *118* (3), 1121–1167.

Clay, Karen, Joshua Lewis, and Edson Severnini, "Canary in a Coal Mine: Infant Mortality, Property Values, and Tradeoffs Associated with Mid-20th Century Air Pollution," Working Paper 22155, National Bureau of Economic Research April 2016.

Cropper, Maureen, Ryna Cui, Sarath Guttikunda, Nate Hultman, Puja Jawahar, Yongjoon Park, Xinlu Yao, and Xiao-Peng Song, "The mortality impacts of current and planned coal-fired power plants in India," *Proceedings of the National Academy of Sciences*, 2021, *118* (5).

Currie, Janet 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.

_ **and Tom Vogl**, "Early-Life Health and Adult Circumstance in Developing Countries," *Annual Review of Economics*, 2013, *5* (1), 1–36.

_ , **Lucas Davis, Michael Greenstone, and Reed Walker**, "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings," *American Economic Review*, February 2015, *105* (2), 678–709.

Dandona, Rakhi, G Anil Kumar, Nathaniel J Henry, Vasna Joshua, Siddarth Ramji, Subodh S Gupta, Deepti Agrawal, Rashmi Kumar, Rakesh Lodha, Matthews Mathai et al., "Subnational mapping of under-5 and neonatal mortality trends in India: the Global Burden of Disease Study 2000–17," *The Lancet*, 2020, *395* (10237), 1640–1658.

der Goltz, Jan Von and Prabhat Barnwal, "Mines: The local wealth and health effects of mineral mining in developing countries," *Journal of Development Economics*, 2019, *139*, 1–16.

Deryugina, Tatyana, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif, "The mortality and medical costs of air pollution: Evidence from changes in wind direction," *American Economic Review*, 2019, *109* (12), 4178–4219.

Deschênes, Olivier and Michael Greenstone, "The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather," *American economic review*, 2007, *97* (1), 354–385.

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

Draxler, Roland R and GD Hess, "An overview of the HYSPLIT_4 modelling system for trajectories," *Australian meteorological magazine*, 1998, *47* (4), 295–308.

_ , **Barbara Stunder, Glenn Rolph, Ariel Stein, and Albion Taylor**, "HYSPLIT4 users's guide," Technical Report 2020.

Druckenmiller, Hannah and Solomon Hsiang, "Accounting for unobservable heterogeneity in cross section using spatial first differences," Technical Report, National Bureau of Economic Research 2018.

Duflo, Esther and Rohini Pande, "Dams," *The Quarterly Journal of Economics*, 2007, *122* (2), 601–646.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik instruments: What, when, why, and how," *American Economic Review*, 2020, *110* (8), 2586–2624.

Gulati, Sumeet, Krithi K Karanth, Nguyet Anh Le, and Frederik Noack, "Human casualties are the dominant cost of human–wildlife conflict in India," *Proceedings of the National Academy of Sciences*, 2021, *118* (8).

Gupta, Aashish and Dean Spears, "Health externalities of India's expansion of coal plants: Evidence from a national panel of 40,000 households," *Journal of Environmental Economics and Management*, 2017, *86*, 262–276.

Guttikunda, Sarath K and Puja Jawahar, "Atmospheric emissions and pollution from the coal-fired thermal power plants in India," *Atmospheric Environment*, 2014, *92*, 449–460.

Heblich, Stephan, Alex Trew, and Yanos Zylberberg, "East Side Story: Historical Pollution and Persistent Neighborhood Sorting," Discussion Paper Series, School of Economics and Finance 201613, School of Economics and Finance, University of St Andrews November 2016.

Henneman, Lucas R. F., Christine Choirat, Cesunica Ivey, Kevin Cummiskey, and Corwin M. Zigler, "Characterizing population exposure to coal emissions sources in the United States using the HyADS model," *Atmospheric Environment*, April 2019, *203*, 271–280.

_ , Irene C. Dedoussi, Joan A. Casey, Christine Choirat, Steven R. H. Barrett, and Corwin M. Zigler, "Comparisons of simple and complex methods for quantifying exposure to individual point source air pollution emissions," *Journal of Exposure Science & Environmental Epidemiology*, July 2021, *31* (4), 654–663. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 4 Primary_atype: Research Publisher: Nature Publishing Group.

Herrnstadt, Evan, Anthony Heyes, Erich Muehlegger, and Soodeh Saberian, "Air pollution and criminal activity: Microgeographic evidence from Chicago," *American Economic Journal: Applied Economics*, 2021, *13* (4), 70–100.

International Energy Agency, *India Energy Outlook 2021* 2021.

Jayachandran, Seema, "Air quality and early-life mortality evidence from Indonesia's wildfires," *Journal of Human resources*, 2009, *44* (4), 916–954.

Kleibergen, Frank and Richard Paap, "Generalized reduced rank tests using the singular value decomposition," *Journal of econometrics*, 2006, *133* (1), 97–126.

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.

Komisarow, Sarah and Emily L Pakhtigian, "Are power plant closures a breath of fresh air? Local air quality and school absences," *Journal of Environmental Economics and Management*, 2022, p. 102569.

Kuminoff, Nicolai V., V. Kerry Smith, and Christopher Timmins, "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets," *Journal of Economic Literature*, December 2013, *51* (4), 1007–62.

Landrigan, Philip J, Richard Fuller, Nereus J R Acosta, Olusoji Adeyi, Robert Arnold, Niladri (Nil) Basu, Abdoulaye Bibi Baldé, Roberto Bertollini, Stephan Bose-O'Reilly, Jo Ivey Boufford, Patrick N Breysse, Thomas Chiles, Chulabhorn Mahidol, Awa M Coll-Seck, Maureen L Cropper, Julius Fobil, Valentin Fuster, Michael Greenstone, Andy Haines, David Hanrahan, David Hunter, Mukesh Khare, Alan Krupnick, Bruce Lanphear, Bindu Lohani, Keith Martin, Karen V Mathiasen, Maureen A McTeer, Christopher J L Murray, Johanita D Ndahimananjara, Frederica Perera, Janez Potočnik, Alexander S Preker, Jairam Ramesh, Johan Rockström, Carlos Salinas, Leona D Samson, Karti Sandilya, Peter D Sly, Kirk R Smith, Achim Steiner, Richard B Stewart, William A Suk, Onno C P van Schayck, Gautam N Yadama, Kandeh Yumkella, and Ma Zhong, "The Lancet Commission on Pollution and Health," *The Lancet*, February 2018, *391* (10119), 462–512.

Lipscomb, Molly, A Mushfiq Mobarak, and Tania Barham, "Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil," *American Economic Journal: Applied Economics*, 2013, *5* (2), 200–231.

Lu, Zifeng and David G Streets, "Increase in NO x emissions from Indian thermal power plants during 1996–2010: unit-based inventories and multisatellite observations," *Environmental Science & Technology*, 2012, *46* (14), 7463–7470.

_ , _ , Benjamin de Foy, and Nickolay A Krotkov, "Ozone Monitoring Instrument observations of interannual increases in SO2 emissions from Indian coal-fired power plants during 2005–2012," *Environmental science & technology*, 2013, *47* (24), 13993–14000.

MacQueen, James et al., "Some methods for classification and analysis of multivariate observations," in "Proceedings of the fifth Berkeley symposium on mathematical statistics and probability," Vol. 1 Oakland, CA, USA 1967, pp. 281–297.

Mays, G. T., R. J. Belles, B. R. Blevins, S. W. Hadley, T. J. Harrison, W. C. Jochem, B. S. Neish, O. A. Omitaomu, and A. N. Rose, "Application of Spatial Data Modeling and Geographical Information Systems (GIS) for Identification of Potential Siting Options for Various Electrical Generation Sources," Technical Report, Oak Ridge National Laboratory December 2011.

Mendelsohn, Robert, William D Nordhaus, and Daigee Shaw, "The impact of global warming on agriculture: a Ricardian analysis," *The American economic review*, 1994, pp. 753–771.

Munshi, Kaivan and Mark Rosenzweig, "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 2016, *106* (1), 46–98.

Murray, Christopher J L, Aleksandr Y Aravkin, Peng Zheng, Cristiana Abbafati, Kaja M Abbas, Mohsen Abbasi-Kangevari, Foad Abd-Allah, Ahmed Abdelalim, Mohammad Abdollahi, Ibrahim Abdollahpour, Kedir Hussein Abegaz, Hassan Abolhassani, Victor Aboyans, Lucas Guimarães Abreu, Michael R M Abrigo, Ahmed Abualhasan, Laith Jamal Abu-Raddad, Abdelrahman I Abushouk, and Maryam Adabi, "Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019," *The Lancet*, October 2020, *396* (10258), 1223–1249.

Olea, José Luis Montiel and Carolin Pflueger, "A robust test for weak instruments," *Journal of Business & Economic Statistics*, 2013, *31* (3), 358–369.

Planning Commission, "Indian Five Year Plan, 1961-66," Policy Document, Planning Commission, Government of India 1961.

Rosales-Rueda, Maria and Margaret Triyana, "The persistent effects of early-life exposure to air pollution evidence from the indonesian forest fires," *Journal of Human Resources*, 2019, *54* (4), 1037–1080.

Shearer, Christine, Matthew-Shah Neha, Lauri Myllyvirta, Aiqun Yu, and Ted Nace, *Boom and bust 2019*, Sierra Club, 2019.

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

Staiger, Douglas O and James H Stock, "Instrumental variables regression with weak instruments," 1994.

Stein, A. F., R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, and F. Ngan, "NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System," *Bulletin of the American Meteorological Society*, December 2015, *96* (12), 2059–2077. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Trippi, Michael H and Susan J Tewalt, "Geographic information system (GIS) representation of coal-bearing areas in India and Bangladesh," Technical Report 2011.

Venkataraman, Chandra, Michael Brauer, Kushal Tibrewal, Pankaj Sadavarte, Qiao Ma, Aaron Cohen, Sreelekha Chaliyakunnel, Joseph Frostad, Zbigniew Klimont, Randall V Martin et al., "Source influence on emission pathways and ambient PM 2.5 pollution over India (2015–2050)," *Atmospheric Chemistry and Physics*, 2018, *18* (11), 8017–8039.