The child health impacts of coal: Evidence from India's coal expansion

Sangita Vyas*

December 28, 2019

Abstract

What are the child health and human capital consequences of India's large coal power expansion? Using variation in local coal capacity within place across cohorts, I find that exposure to a median-sized coal plant at birth is associated with a 0.1 standard deviation child height deficit. Supporting air pollution as a channel, effects are larger among children living closer to coal plants. Changes in coal capacity do not predict changes in other child observables, or luminosity, which indicates electricity coverage and economic development. Effects are similar for rich and poor, but rich households tend to live closer to coal plants.

*University of Texas at Austin, sangita.vyas@utexas.edu. 2225 Speedway, Austin, TX 78712. This paper benefited from helpful comments received at PAA 2019, NEUDC 2019, and the UT Austin Applied Micro Lunch Seminar. I thank Sam Arenberg, Diane Coffey, Mike Geruso, Aashish Gupta, Nazar Khalid, Kevin Kuruc, Leigh Linden, Melissa Lopalo, Seth Neller, Gerald Oettinger, Dean Spears, Nikhil Srivastav, Steve Trejo, and Tom Vogl, for useful conversations and feedback. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1610403. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. This research was also supported by grant, P2CHD042849, Population Research Center, awarded to the Population Research Center at The University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

1 Introduction

Air pollution exposure has important consequences for public health. A large literature in economics documents the health effects of early-life exposure to air pollution (see Currie et al. (2014) for a review). However, much of this literature focuses on developed countries, which are more likely to have high-quality data on health and air quality. The health impacts of pollution in developing countries are important to understand, though, because pollution levels are often much higher, and infant health is more fragile. For instance, Arceo, Hanna and Oliva (2016) explore the effect of air pollution on infant mortality using thermal inversions over Mexico City as an instrument for pollution spikes, and for some pollutants find larger health effects than studies in developed countries. Similarly, Tanaka (2015) studies an environmental regulation in China that limited industrial emissions, and finds reductions in infant mortality post policy change that are substantially larger than effect sizes found in developed countries (Knittel, Miller and Sanders, 2016; Currie and Neidell, 2005; Chay and Greenstone, 2003). This paper studies the Indian context, and estimates the effect on child health of growing up near a coal plant, which exposes children to harmful air pollutants including particulate matter, nitrogen dioxide, and sulfur dioxide.

Among developing countries, India is one of the largest consumers of coal. Over the past decade, coal plant capacity has increased rapidly, reaching about three-quarters of India's total electricity generation in 2016 (see Figure 1). Because coal plants in India often do not meet emissions regulations (Bhati et al., 2015), the increases in air pollution associated with India's expansion of coal plant capacity may present large negative health externalities. However, despite the importance of this question, no prior paper investigates the consequences of India's coal plants for human capital.

Between 2010 and 2016, the period over which children in this dataset were born, coal plant capacity more than doubled in India. I exploit variation in the timing and geography of these coal plant capacity additions to identify the effect of coal plant exposure on child height in a difference-in-differences framework. I have the precise location, commissioning date, and generating capacity, for every new generation unit that went online during this period of time. I link geocoded villages and urban areas surveyed in India's most recent Demographic and Health Survey (DHS) to the dataset on coal plants based on proximity. My strategy uses variation in exposure within places over time, controlling for cohort differences that are common across all places.

As a measure of child health, I use child height, an important economic variable that predicts cognitive development, educational attainment, and adult earnings (Case and Paxson, 2008). Child height has been identified as a summary measure of net nutrition, indicating both the disease and nutritional burden in childhood (Bozzoli, Deaton and Quintana-Domeque, 2009). Air pollution is plausibly linked to child height through the established impact of pollution on birth weight (Rangel and Vogl, 2018; Currie and Walker, 2011), and the relationship between birth weight and stature in childhood (Binkin et al., 1988).

I find that an additional median-sized coal plant (in terms of capacity) is associated with a height deficit of about 0.1 standard deviations. This effect is about one-third of the within place height gap between children of illiterate versus literate mothers, about equal to the within place difference in height between children born outside versus within an institutional facility, and about equivalent to the expected decrease in height associated with a one-third standard deviation reduction in wealth. Because the nutrition that contributes to linear growth also contributes to cognitive development (Richards et al., 2002), child height is a marker of earnings later in life. Therefore, these findings also suggest that exposure to the pollution from coal plants may have meaningful and enduring negative consequences for India's economy.

A battery of mechanism and falsification checks provide support for an interpretation of a causal effect of coal plants on child height. The empirical patterns are consistent with the underlying mechanism of air pollution: coal plant capacity expansions are associated with increases in air pollution, and the effect on child height attenuates in distance from the coal plant. Effects are also robust to falsification tests. Changes in coal plant capacity do not similarly predict changes in other birth characteristics related to child height, nor do they predict changes in light output at night, an indicator of electricity coverage and economic development. This may be because the electricity generated from these plants is used to power areas both near and far, facilitated by India's large centralized grid. Two further tests alleviate endogeneity concerns related to where new plants locate. Places near new coal plants do not have differential height trends compared to places that never end up near coal plants. Similarly, capacity additions sometimes occurred through expansions to existing plants and sometimes through entirely new coal plants, and the main effects hold when using only the variation from expansions. Finally, the effects are not explained by rich households migrating away from coal plants, or poor households migrating towards them.

Heterogeneity analyses find similar effects for children of high and low socio-economic status (SES) households. However, paradoxically, high SES households in India are more likely than low SES households to live close to coal plants. Because the effect of a coal plant attenuates in distance from the plant, distance may be an important interaction in heterogeneity analyses. Although the dataset is not powered to detect statistically different effects by SES and distance from a coal plant, the patterns of effects and geography suggest that, after distance is accounted for, coal plants may be more harmful for children of low

SES compared to high SES households.

A growing literature on developing countries documents the child health impacts associated with polluting activities, such as agricultural fires (Pullabhotla, 2019; Singh et al., 2019; Rangel and Vogl, 2018) and forest fires (Jayachandran, 2009; Frankenberg, McKee and Thomas, 2005). These activities are easily observable, and are useful signals of air pollution when air quality monitoring is poor or non-existent. This paper studies electricity generation from coal, an important source of air pollution in developing countries that has potentially large population health impacts. In the developed-country context, Clay, Lewis and Severnini (2016) investigate an expansion in coal plants in the United States in the early 20th century. Comparing outcomes in counties within 30 miles of new coal plants to those in counties within 30 to 90 miles in a fixed effects framework, they find that increased coal consumption led to higher infant mortality rates. Because the pollutants generated from burning coal have different properties compared to other pollution sources, and because Indian coal has particularly high ash content, investigating the health impacts associated with air pollution from coal plants in India is important to study directly. Gupta and Spears (2017) investigate the respiratory health effects of the same Indian coal plant expansion that I study here. Taking advantage of a large panel dataset of Indian households, they show that reported coughs decreased by less in places in which coal plant exposure increased by more between 2005 and 2012. The effect of India's coal plant expansion on other indicators of child health is an open area of investigation.

This article makes several contributions to the literature. First, I study the impact of coal plant exposure on child height, an important health outcome that has received little attention in the air pollution literature. One potential biological mechanism linking air pollution exposure to child height operates through pregnant women's health. Exposure to particulate matter may cause infections in pregnant women that exacerbate pulmonary or placental inflammation and hinder the exchange of oxygen and nutrients between the placenta and the fetus (van den Hooven et al., 2012). This may affect fetal growth, birth weight, and stature in childhood. A second possible mechanism operates through the effect of air pollution on the respiratory health of young children. Air pollution increases the incidence of respiratory infections among children (Pope III et al., 2011), and the immune response brought about to fight disease plausibly diverts scarce nutrients away from physical development (Crimmins and Finch, 2006).

Second, I focus on the developing country context, where understanding the health impacts of coal is especially pressing because coal still comprises a large fraction of electricity generation. Importantly, the effects estimated by Clay, Lewis and Severnini (2016) in the U.S. may not be applicable for developing countries like India for several reasons. The coal plants under study in the U.S. are much smaller than the coal plants that are currently becoming operational in India and other developing countries, and the quality of coal in India and other developing countries is likely to be worse. For these reasons, the associated pollution levels in India are likely to be higher. If the health effects of particulate pollution are not linear, then the effect sizes estimated in the U.S. may not apply in contexts where the associated pollution levels are higher. Additionally, infant health in India is particularly fragile due to exposure to open defecation (Spears, 2018) and poor maternal nutrition (Coffey, 2015), among other risks, and the effect of air pollution in this context may be different compared to places where baseline health is more robust. Despite clear policy importance, well-identified estimates of the effect of coal plant exposure on health outcomes which have long-term implications does not exist from developing countries, and this study seeks to fill this gap.

Third, this paper contributes to a literature linking the environment and economic development. Jayachandran (2009) finds more severe effects from exposure to wildfires in poor areas compared to rich areas. I do not find statistically different effects between children of rich and poor households. However, in contrast to other contexts (Chay and Greenstone, 2003; Currie and Walker, 2011), I find that high SES households are more likely to live closer to coal plants than low SES households, despite greater exposure to air pollution. Although the dataset is not powered to estimate statistically different effects by SES and distance, these findings highlight complex distributional consequences from electricity generation from coal plants.

The paper proceeds as follows. Section 2 summarizes the datasets used for this analysis, and Section 3 describes the identification strategy. Section 4 discusses the main results and patterns that are consistent with the mechanism of air pollution. Section 5 runs through a battery of falsification tests and robustness checks. Section 6 discusses extensions to the main analysis including heterogeneous effects and an exploration of exposure timing. Section 7 concludes.

2 Data

The main analyses in this paper use data on child health from India's Demographic and Health Survey (DHS) 2015-2016 and a dataset on power plant openings and capacity in India from 1922 to 2018. In mechanism and falsification tests, I supplement these with data on $PM_{2.5}$, particulate matter smaller than 2.5 microns in diameter, estimated from satellite measurements of aerosol optical depth, and light output at night, estimated from nightly photographs taken of the Earth from space.

Data on child height, the dependent variable of interest, and other characteristics of children and their mothers come from India's DHS, which interviewed a nationally-representative sample of women of reproductive age between January 2015 and December 2016. Surveyors measured the heights of all children of surveyed women under the age of five. As is common in the literature on child height, height is standardized using the mean and standard deviation, by age and sex, of a healthy reference population identified by the World Health Organization (WHO et al., 2006).

The independent variable of interest is coal plant capacity. Data on the openings, closures, and plant capacity of all power plants in India are obtained from the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector. I match each coal plant in this dataset to urban blocks and rural villages, hereafter called villages for simplicity, surveyed in the DHS based on proximity. Following Clay et al. (2016), villages that are within 50 km of a coal plant are considered exposed to the plant. Villages that are more than 50 km from all coal plants are considered not exposed.[1] [2] The 50 km cutoff is also validated in an analysis that tests effects by distance from the plant, discussed in Section 4.4.

The third dataset consists of district-month-year estimates of $PM_{2.5}$ from 2010 to 2015 from Dey et al. (2012). Because India lacks ground-based pollution measurements at a spatial resolution sufficient for my study design,[3] I use data estimated from satellite measurements. Specifically, estimates of $PM_{2.5}$ are generated from aerosol optical depth data collected using the Multiangle Imaging SpecroRadiometer (MISR) V22 product, at 17.6 km $\times$ 17.6 km spatial resolution. Aerosol optical depth indicates how much direct sunlight is scattered or absorbed by aerosol products in the atmosphere. Estimates of $PM_{2.5}$ were constructed using chemical transport model simulations that included aerosol optical depth, emissions, and meteorological factors like temperature, relative humidity, and precipitation. The estimates and methodology used for generating them are presented in detail in Dey et al. (2012). I match this data on air quality to district-month-year measurements of coal plant capacity in order to study whether increases in coal plant capacity are associated with increases in air pollution within districts over time.[4] It is important to note, however, that districts

[1] In alternative models, villages that are more than 50 km from all coal plants, and between 50 and 100 km from at least one coal plant, are the unexposed group used in analyses, and villages beyond 100 km from all coal plants are dropped.

[2] In order to maintain respondent confidentiality, The DHS Program randomly displaces the GPS latitude and longitude positions for all surveys. Urban clusters are displaced between zero and two km from the actual location. Rural clusters are displaced between zero and five km, with one percent of rural clusters displaced between zero and ten km. This displacement technique introduces measurement error to the exposure variable. Thus, the true effect of coal plant exposure on height may be larger than estimated.

[3] There are, in fact, no air quality monitors in rural India.

[4] GPS coordinates and district are available for all coal plants. I construct coal plant capacity at the

vary substantially in area, with the smallest district having an area of about 9 km^2, and the largest district having an area of about 45,500 km^2. Considering the variation in district areas relative to the area considered exposed to coal plants, along with the fact that coal plants can be placed on borders of districts thereby exposing villages in multiple districts, it is likely that the magnitude of the effect of district-level coal plant capacity on district-level $PM_{2.5}$ is attenuated and does not reflect the true effect size.

The final dataset contains estimates of light output at night for each month-year from 1993 to 2013 for all villages in India listed in the 2001 Indian Census. This dataset was collected and made publicly available by The World Bank and Min (2017). Light output is estimated from photographs taken of the Earth from space each night by the Defense Meteorological Satellite Program. From these raster images, researchers extracted daily light output values for the pixels corresponding to each village's approximate latitude and longitude. They then filtered and aggregated the data to the month level using a methodology described in greater detail in The World Bank and Min (2017). Luminosity data for 2001 Census villages cannot be matched directly to urban blocks and rural villages in the DHS (which I call villages for simplicity), because DHS locations are randomly displaced to ensure respondent confidentiality. Most DHS villages are displaced by less than 5 km. I therefore draw a circle with a radius of 5 km around each DHS village, and compute the mean of median luminosity values for all 2001 Census villages within the circle. Means are computed in this way for each village-month-year. Luminosity data has been used as a measure of electricity coverage over time (Min, 2015), and as an indicator for economic development in places that have poor quality statistical systems or for geographic levels that are not available from other datasets (Chen and Nordhaus, 2011; Henderson, Storeygard and Weil, 2011).

Figure 2 shows all coal plants and villages, both exposed and unexposed, in the matched dataset of coal plants and villages visited in the DHS. Although coal plants are spread all across India, there is a higher concentration of them in eastern India, and a lower concentration of them in western India. This figure also demonstrates the representativeness of the data; the DHS visited all parts of the country.

district level by summing capacity from all coal plants in the district, for each month-year. I merge this dataset with the district-month-year pollution dataset to generate a district-level dataset on coal plant capacity and $PM_{2.5}$ over time.

3 Econometric framework

I use a difference-in-differences estimation strategy to identify the effect of coal plant exposure on child height. I include place and cohort fixed effects to control for variation in capacity and child height over space and secular changes across cohorts. Because the DHS measured the heights of children at different ages, and child height deficits evolve over time, I also include age in months-by-sex fixed effects.[5] This analysis answers the question: are children born at times when coal plant capacity in the village is higher than average shorter than average, controlling for trends across cohorts that are common to all villages?

I estimate regressions of the following form:

$$height_{ihvt} = \beta coal_{vt} + \boldsymbol{B_{ihvt}\delta} + \boldsymbol{H_{hvt}\gamma} + \boldsymbol{\alpha_t} + \boldsymbol{\mu_v} + \epsilon_{ihvt} \tag{1}$$

where i indexes individual children, h indexes households, v indexes villages (which, for simplicity, refers to urban blocks and rural villages surveyed in the DHS), and t indexes birth cohort in month-years (e.g. March 2014). *height* is the height-for-age z-score of the child, measured at the time of the survey. *coal* is the total capacity, or total number of units, within 50 km in the month of birth. In regressions using capacity, this variable has been rescaled so that one unit of capacity represents one median-sized coal plant, which has 1,000 megawatts (MW) in this data. Coal plant expansions occur either by the construction of new coal plants, or by adding additional generation units to existing coal plants. I use the number of generation units installed as the independent variable of interest in alternative models. α represents cohort fixed effects for the month-year of birth, and μ represents village fixed effects. B represents birth characteristics, including a full set of age-by-sex indicators, mother's age at birth, birth order, multiple birth, institutional delivery, and c-section delivery.[6] In some models, I also include whether or not the mother took iron supplements or anti-helmintics during pregnancy, variables which were only available for the youngest child under five. H represents mother and household characteristics, including mother's height, religion, caste, literacy, household open defecation, and use of solid fuels for cooking, variables that are indicative of SES. It is important to note that these household variables are only observed at the time of the survey, and may not represent the household environment at the time the child was born, or in early life. Most of these variables, however,

[5]In India, and in other developing countries where environmental risks such as open defecation are particularly severe, average height-for-age is decreasing in age because height reflects the accumulating impact of early-life health insults on a child's growth (Victora et al., 2010).

[6]Fixed effects for the month-year of birth and for age-by-sex, where age (in months) is at the time of measurement, can be identified separately because the DHS collected data over a period of two years. This means that there exist observations of children measured at the same age (in months), born into different cohorts (in month-years).

are likely to be highly correlated over time.

3.1 Identifying variation

Figure 3 shows a histogram of the identifying variation, the change in coal plant capacity within villages from February 2010 to November 2016, which is the period over which children in the dataset were born. As mentioned earlier, for simplicity I use the term 'village' to refer to urban blocks and rural villages surveyed in the DHS. Most villages in the dataset were never exposed to a coal plant over this period, so the median village experienced no increase in coal plant capacity. Among villages that were exposed to coal plants, the median village experienced an increase in capacity of 960 MW, or approximately one extra median-sized coal plant. A small fraction of villages experienced a decrease in coal plant capacity, but these decreases were small in magnitude and in frequency, relative to the increases. The dashed line near 1,000 MW corresponds to the increase in exposure associated with a plant opening with median 2016 capacity. Figure 3 also shows that the distribution of the change in coal plant capacity has a long right tail: the 75th percentile village experienced an increase in capacity of 1,400 MW, and the village that had the greatest increase in coal capacity saw an increase of 10,580 MW. In robustness checks presented in Section 5.5, I test whether the results are sensitive to outliers.

An important feature to note is that coal plant capacity exposure within villages is highly correlated over time. Figure 4 shows village-level exposure for two selected villages, one that experienced the 25th percentile increase in capacity, and one that experienced the 75th percentile increase. Notably, exposure remains constant for several years in each village, before increasing. This high correlation in exposure over time complicates analyses to determine the specific months of child development that are most sensitive to air pollution exposure. I therefore use exposure in the month of birth *a priori* because it represents a period in which environmental risks are important for child health and development (Currie and Vogl, 2013). Section 6.2 explores exposure timing in greater detail.

3.2 Identifying assumptions

Although this strategy permits villages that are exposed to coal plant capacity to be different in terms of levels from villages that are not, it relies on the assumption that additions in capacity are exogenous conditional on fixed effects and control variables. Put differently, places in which coal plant capacity increased would have trended in parallel, had they not gotten increases in capacity, to places in which capacity did not increase. This assumption would be violated if, for instance, the expansion of a coal plant brought more labor market opportunities to local residents and made households richer. The bias that this could introduce to an estimate of the effect of coal plants on child health is not directly apparent since there could be competing income and substitution effects from increased wages (Miller and Urdinola, 2010). Another type of violation would arise if, for instance, poorer households moved near new coal plants because these areas became more affordable due to a degradation in location quality. This particular violation would bias impacts in the direction of a deterioration in child health, and would produce more negative effects than can actually be attributed to changes in capacity. In fact, I show in Section 5.4 that migration of women and children is uncommon, and is not driving the results.

I indirectly test the identification assumption in multiple ways in Sections 4 and 5. First, I show evidence that the effect appears to operate through air pollution: coal plant capacity expansions are associated with increases in air pollution, and capacity expansions that are farther away have weaker health effects. Then, I test whether coal plant capacity predicts other observable birth and village characteristics that may be related to child height, and show that it does not. I also analyze pre-trends in child height in places that became newly exposed to coal plants after the DHS completed data collection, compared to places that remained unexposed. Height trends were similar in both types of villages. Finally, I present evidence suggesting that the results are not driven by migration, and that in fact, richer households tend to live nearer to coal plants.

4 Results

4.1 Descriptive statistics

Table 1 shows summary statistics for the children in the data, stratified based on living in a village exposed to coal plant capacity. As mentioned earlier, for simplicity I use the term 'village' to refer to urban blocks and rural villages surveyed in the DHS. A notable feature of the children in this data is that, irrespective of their exposure to coal plants, they are very short based on international standards. In both types of villages, children are about 1.48 standard deviations shorter than the healthy reference population. On some measures that are important for child health, children in exposed villages do better than children in unexposed villages. On other measures, they do worse. For example, children in exposed villages are more likely to have a literate mother, are less likely to live in a rural area, and are less likely to live in a household that defecates in the open and uses solid fuels. On the other hand, children in exposed villages are less likely to have been born in a hospital or public health center, have shorter mothers, and were less likely to have started breastfeeding within the first day. Although these differences are not ideal, it is important to note that they do not *a priori* invalidate the identification assumption because the main analyses identify effects from variation within villages over time, and therefore control for both observed and unobserved fixed village characteristics.

4.2 Changes in coal capacity predict changes in air pollution

Table A1 verifies that coal plant capacity is associated with higher levels of air pollution, as measured by $\mathrm{PM}_{2.5}$, at the district-month-year level.[7] In models with and without fixed effects for district and month-year, an extra gigawatt in coal plant capacity is statistically significantly associated with between one and two $\mu\mathrm{g/m}^3$ more $\mathrm{PM}_{2.5}$.[8] It is important to note, however, that these estimates are likely to be highly attenuated because many districts are larger than 100 km in diameter, and therefore district-level $\mathrm{PM}_{2.5}$ estimates average over spaces that likely include exposed and unexposed regions. The true effect of coal plant capacity on $\mathrm{PM}_{2.5}$ in areas near coal plants is likely to be much larger than these estimates. Nevertheless, this table establishes the presence of a statistical relationship, and provides evidence in support of air pollution as a mechanism.

4.3 Main result: coal plants predict child height

Table 2 shows the main results of the analysis. In all models in this table, the dependent variable is height-for-age z-score and the sample consists of all children with measured height. In panel A, the independent variable is coal plant capacity. A one unit increase in capacity corresponds to an increase of 1,000 MW, or 1 gigwatt (GW), the size of the median coal plant in the dataset. In panel B, the independent variable is the number of coal generation units. The median plant in the data has three generation units.

There are at least three important features to note in this table. First, although the effect sizes in panels A and B look quite different at first glance, they are consistent with each other. Multiplying the effect sizes in panel B by three (the number of generation units in the median-sized plant) produces effect sizes that are very similar in magnitude to those shown in panel A. Second, it is unsurprising that the coefficients in panel B are noisier than those in panel A. This is because the number of generation units in a coal plant is a noisier predictor of what we believe is harmful for child height, the pollution generated from burning coal. Finally, each column in this table corresponds to a regression with a slightly different

[7]Outside of the economics literature, this data has been used in a public health study by Spears et al. (2019); they do not observe or study coal plants.

[8]Regressions are of the form: $PM2.5_{dt} = \beta capacity_{dt} + \boldsymbol{\alpha_t} + \boldsymbol{\rho_d} + \epsilon_{dt}$, where t represents month-years, d districts, α month-year fixed effects, and ρ district fixed effects.

specification. Notably, as I add more controls to the regression, the coefficient on capacity, and on units, stays relatively constant. Across all models, the results are very quantitatively similar: an additional median-sized coal plant is associated with a height deficit of about 0.1 standard deviations.

This is a large and economically important effect. Child height is correlated with socioeconomic status, and this effect size represents a relatively large gap in status. For instance, the main effect is about one-third of the within place height gap between children of illiterate versus literate mothers, about equal to the within place difference in height between children born outside versus within an institutional facility, and about equivalent to the expected decrease in height associated with a one-third standard deviation reduction in wealth. Moreover, these findings are consistent with the effects on height from exposure to poor sanitation. The effect of an extra median-sized coal plant on child height is about the same as the effect of reducing local area open defecation by about 25 percentage points (Gertler et al., 2015).

Column 1 shows results from a model that includes age-by-sex, month-by-year of birth (cohort), and village fixed effects. Column 2 adds birth characteristics, including mom's age at birth, birth order, multiple birth, institutional delivery, and c-section delivery. Column 3 adds mother and household characteristics, including mother's height, religion, caste, literacy, household open defecation, and solid fuel use. In column 4, I allow for geography-specific cohort trends by including latitude-by-cohort and longitude-by-cohort trends.[9] Geography-specific cohort trends allow different parts of India to have different cohort trends. Finally, in column 5, I exclude the cohort trends, and include iron supplementation and anti-helmintics during pregnancy. Since these variables were only available for the mother's last birth, these regressions have smaller sample sizes than the other models. Across models, the effect of coal plant capacity remains very stable. The fact that different models produce similar results provides evidence that none of the potential confounders that these specifications address are driving the observed height results.

[9]Latitude, longitude, and time in month-years are continuous variables in these interactions. An alternative test for differential cohort trends allows states to have differential cohort trends. In a model that builds off of the specification in column 3 by adding state-by-cohort fixed effects, the coefficient on *capacity* is -0.0510 (standard error = 0.0411). This coefficient is roughly half the size of the coefficients in other models, and is no longer statistically significant. An attenuated coefficient is expected since there is a higher density of coal plants in particular states, and since it is possible that children born outside of the 50 km radius of a coal plant, but still living in the state, are also affected.

4.4 Effects are larger for children nearer to a coal plant

If coal plant capacity affects the heights of children through air pollution, an effect that attenuates in distance would be consistent with this mechanism. I test this hypothesis in Figure 5. This figure plots the regression coefficients from a single regression of height-for-age on capacity within different distance bins, from less than 20 km, to between 60 to 70 km.[10] Coefficients are produced from the following regression:

$$\begin{aligned}
height_{ivt} = &\beta_1 capacitywithin20km_{vt} + \beta_2 capacity20-30km_{vt} + \\
&\beta_3 capacity30-40km_{vt} + \beta_4 capacity40-50km_{vt} + \\
&\beta_5 capacity50-60km_{vt} + \beta_6 capacity60-70km_{vt} + \\
&\boldsymbol{B_{ivt}\delta} + \boldsymbol{\alpha_t} + \boldsymbol{\mu_v} + \epsilon_{ivt}
\end{aligned} \tag{2}$$

where $height$, i, v, t, α and μ are as described in Equation 1. The vector of birth characteristics, B, represents indicators for age-by-sex categories.

Except for the first distance bin, the effect size of coal plant capacity becomes closer to zero as distance from the coal plant increases.[11] The effects of capacity within bins less than 40 km away are statistically different from zero at the 5 percent level, the p-value on the 40 to 50 km bin is 0.14, and beyond 50 km, the p-values on capacity effects are greater than 0.30. This finding lends credence to the 50 km cutoff used in the main results, since exposure beyond this distance does not statistically significantly influence child height. Table A2 tests whether the size of the effect is decreasing in distance by interacting capacity and distance. The interaction term is statistically significant, and the magnitude indicates that the size of the effect decreases by 0.017 for every 10 km.[12] In summary, the patterns presented in this section are consistent with an effect of coal plants on child height through the mechanism of air pollution.

[10]Within 10 km, and between 10 and 20 km, are combined into one single bin because of small sample sizes.

[11]The coefficient on the first distance bin is slightly closer to zero than the coefficient on the second distance bin, but a test of the hypothesis that these two coefficients are the same cannot be rejected (F-statistic = 1.00, p-value = 0.32).

[12]Given these coefficients, the effect of capacity reaches zero at about 80 km distance from the plant. This is considerably larger than the 50 km distance cutoff used in the main analysis. These results do not contradict the main analysis, however, because the distance variable used in Table A2 is a weighted average of all coal plants within 70 km of the village, where weights are the fraction of total capacity that the coal plant contributes for that village in that month.

5 Falsification tests and robustness checks

5.1 Changes in observable child characteristics

If increases in coal plant capacity were occurring at the same time as other changes that matter for child health, differently between exposed and non-exposed villages, the effects found in Section 4 would be biased. Table 3 presents evidence suggesting that other coincident changes were not taking place. This table shows regression results using the same specification as in Table 2, column 1, but replacing height-for-age with other variables that are important for child health: mother's age at birth, birth order, multiple birth, institutional delivery, c-section delivery, early initiation of breastfeeding, iron supplementation, and anti-helmintic drugs. None of the estimates displayed in Table 3 are significant, indicating that coal plant capacity does not predict changes along any of these other dimensions that are important for child health.[13] Although this does not rule out the possibility that other changes relevant for child health occurred at the same time that capacity increased, these results provide suggestive evidence that the main findings are not driven through any other channels.

5.2 Changes in electricity and economic development

The DHS data cannot be used to study how household wealth changed over time relative to coal plant capacity because the survey only measured household characteristics once, at the time of the survey. In order to test whether changes in coal plant capacity predict changes in infrastructure and economic factors that may be relevant for child height, I merge in village-level data on light output at night made publicly available by The World Bank and Min (2017). Table 4 tests whether coal plant capacity predicts night lights.[14] Observations in this analysis are village-month-years. Unsurprisingly, coal plant capacity statistically predicts

[13] Table A3 shows statistically significant associations of height-for-age z-score with each of these characteristics. In contrast to other settings where more clinically complicated deliveries that are more likely to result in shorter and less healthy children may be carried out by c-section, or take place in an institutional facility, in India, c-section and institutional deliveries are associated with taller children, delivered by wealthier and healthier moms. Mom's age at birth, iron supplementation during pregnancy, drugs for intestinal parasites during pregnancy, mom's literacy, mom's height, and household wealth quintile, are all positively and statistically significantly associated with a child's height-for-age z-score. Multiple birth and birth order are associated with shorter children. Unexpectedly, early initiation of breastfeeding appears to be negatively associated with height-for-age z-score, although it is marginally statistically significant.

[14] Village-level regressions are of the form: $night\ lights_{vt} = \beta capacity_{vt} + \boldsymbol{\alpha_t} + \boldsymbol{\rho_v} + \epsilon_{vt}$, where t represents month-years, v villages, α month-year fixed effects, and ρ village fixed effects. For each village-month-year, *night lights* is the mean of median monthly luminosity across all 2001 Census villages that are located within a 5 km radius of the DHS village. See Section 2 for further description of the night lights dataset.

light output at night in the model that does not include geographic or temporal fixed effects; higher coal plant capacity is associated with more light output at night. This is because villages that are exposed to coal plants are more urban and have greater electricity coverage than villages that are not exposed to coal plants. When we include village fixed effects in column 2, however, this relationship substantially attenuates and becomes statistically insignificant. Within villages, changes over time in coal plant capacity are not statistically associated with changes in light output at night. Column 3, the model that most closely approximates the estimation strategy in the main analysis, includes both village and month-year fixed effects, and the point estimate attenuates even more. This implies that village luminosity is not greater than average when the village has higher capacity than its average, controlling for trends over time in luminosity and coal plant capacity. Although perhaps surprising, this finding makes sense in the Indian context. India's electricity grid is well-connected, making it possible to transmit power from a plant to meet demand somewhere else.

Chen and Nordhaus (2011) and Henderson, Storeygard and Weil (2011) investigate night lights as a proxy for standard measures of output and find that they provide informational value in countries with low-quality statistical systems and sub-national regions for which no other data are available. If changes in luminosity are indicative of economic development, Table 4 provides evidence that changes in coal plant capacity do not predict differential development patterns in exposed villages compared to unexposed villages. Even if night lights are not a robust predictor of economic development, this analysis documents that changes in coal capacity do not predict changes in electricity coverage. This suggests that the electricity generated from coal plants in India contributes to increasing electricity coverage in unexposed areas just as much as in exposed areas. To summarize, these patterns suggest that economic growth and electricity coverage may not be confounding variables in the analysis of coal plants and child height.[15]

5.3 Plant placement and pre-trends

This section addresses endogeneity concerns around plant placement. In a preliminary test, I only use the variation in capacity that arose from expansions to existing plants only, and

[15] Table A4 includes night lights as a covariate in the main analysis. Column 1 of this table repeats the main results presented in column 1, Table 2. Column 2 implements the same regression as in column 1, but on the sample of children for which night lights data are available. The sample size in column 2 is less than half the sample size in column 1. Column 3 uses the same sample as column 2, but adds night lights as an explanatory variable. The coefficients on coal plant capacity in columns 2 and 3 are very similar in magnitude and statistical significance, and the coefficient on night lights in column 3 is of a very small magnitude and is statistically insignificant. This provides further evidence that the main results are not biased by differential changes in economic development or electricity coverage between exposed and unexposed villages.

drop variation from the opening of entirely new coal plants. The results of this analysis are shown in Table A5, column 4, and are very similar to the main results.

I also test pre-trends in child height in villages that became newly exposed to coal plant capacity after the DHS completed data collection, compared to villages that did not get exposed to future coal plants. It is important to note, here, that the heights of children are measured at the time of the survey. Therefore, children born before a coal plant starts could still be exposed to the coal plant later in their lives. Consider, for example, a child born in January 2012, at a time when coal plant capacity in her village is zero. In January 2013, a coal plant opens up nearby, and capacity in the village increases to 1,000 MW. Then, in January 2015, the DHS team visits the child's village, interviews her mom, and measures her height and the heights of all other siblings under the age of five. Although this child was not exposed to coal plant capacity in her month of birth, she was exposed to it from the age of 12 months until her height was measured in January 2015. If exposure to coal plant capacity beyond the month of birth is important for child health, this child's height may be shorter than it would have been had she not been exposed starting at 12 months. Section 6.2 cannot rule out that exposure until 24 months of age is relevant for child height.

Because of the uncertainty around the relevant period of exposure, I study height trends in villages that became newly exposed to coal plants after the DHS completed data collection in December 2016. I compare villages that were never exposed to coal plants between 2010 and December 2016, but became exposed to a coal plant after December 2016, to villages that were never exposed even after December 2016. Table 5 shows the results of this analysis. The regression equation used for this test is:

$$height_{ivt} = \beta futureplant_v \times time_{ivt} + \eta time_{ivt} + \boldsymbol{B_{ivt}\delta} + \boldsymbol{\mu_v} + \epsilon_{ivt} \tag{3}$$

where $height$, i, v, t, and μ are as described in Equation 1. $future\ plant$ is a dummy variable that varies at the village level, and indicates that the village became exposed to a coal plant after December 2016. β is the coefficient of interest, and indicates whether the heights of children born in villages that became exposed to future coal plants trended differently over time compared to villages that never received a future coal plant. $time$ is a continuous variable, indicating month-year of birth. The vector of birth characteristics, B, represents indicators for age-by-sex categories. An economically small and insignificant β would indicate that the heights of children born from 2010 to 2016 in villages that got coal plants after 2016 did not have a statistically different cohort trend from the heights of children born in villages that never got a coal plant after 2016. Since only children born in villages that are unexposed by December 2016 are included in this analysis, it does not suffer from including any potentially partially treated children.

Table 5, panel A, shows the result of this analysis. $\hat{\beta}$ is indeed small in magnitude, and not significantly different from zero. As expected, $\hat{\eta}$ is positive and significant at the 10 percent level, indicating that children are getting slightly taller across cohorts. Panel B shows the main effect estimated from Table 2, column 1, for comparison. The differential cohort trend is several orders of magnitude smaller than the effect I estimate from exposure to coal plant capacity, and goes in the opposite direction. To summarize, height trends appear to be similar in places that received coal plants compared to places that did not. This alleviates concerns around differential trends in places that received expansions in coal capacity compared to places that did not.

5.4 Migration

If changes in coal plant capacity motivated certain households to move away, and others to move in, the results in the main analysis would be biased. For instance, the decline in air quality may have reduced locational quality, motivating richer households with potentially healthier children to move away and poorer households, who are more likely to have children with weaker health, to move in. On the other hand, the prospect of jobs, electricity, and growth may have motivated more enterprising households to move in. Because the DHS interviews mothers that live in the village at the time the DHS team visits the village, households that may have moved away as a result of the expansion in coal plant capacity are already absent from the main analysis. This section documents facts suggesting that migration of children is not very common, and migration of households to places near coal plants is not likely to bias the results.

Only 8% of children under the age of five in the DHS were born in a location different from the village in which their mother was interviewed. This alone suggests that a mother is unlikely to migrate with her children after her children are born. Moreover, being born in a different location from the interview village does not statistically predict child height, living in an exposed versus unexposed village at the time of the survey, or coal capacity exposure at the time of the survey.[16]

Columns 2 and 3 of Table A5 modify the main analysis to test for the importance of migration. Column 2 restricts the sample to children who were born in the interview village to ensure that each child is assigned the correct level of exposure. The results here are very

[16]Children born in a different location from the DHS interview village are 0.010 (s.e. 0.029) height-for-age z-scores taller, 0.054 (s.e. 1.9) percentage points more likely to be subsequently exposed to coal plants, and are exposed to 0.012 (s.e. 0.028) GW more coal plant capacity at the time of the survey. None of these differences are statistically significant.

similar to the main results. In column 3, I only include children of mothers who have lived in the survey village for at least the six years leading up to the survey. This model removes mothers who may have moved to exposed villages as a result of the changes brought about by increases in coal plant capacity. The sample of included mothers, and their children, are those that lived in the survey village throughout the period of study. Again, the results of this model are very similar to the main results. These findings suggest that changes in coal plant capacity in exposed villages did not induce migration of households with particularly healthy or less healthy children any differently than the movement of households in unexposed villages.

5.5 Other robustness checks

The control group in the main regression results shown in Table 2 consists of all households located more than 50 km from all coal plants. In an alternative analysis shown in Table A6, I follow the strategy employed by Clay, Lewis and Severnini (2016) and Currie and Walker (2011) to construct an alternative control group consisting of children in households that are located more than 50 km from all coal plants, and between 50 and 100 km from at least one coal plant. Summary statistics for this group of children are shown in Table A7. The restriction of the control group in this way associates each child with a nearest coal plant. Therefore, although Table 2 clusters errors at the district level, Table A6 clusters errors at the level of the nearest plant.[17] Comparing these two tables shows that the results are robust to an alternative specification of the control group, and alternative methods of clustering errors (Cameron and Miller, 2015).

There is an ongoing debate in the economics and epidemiology literatures on the shape of the concentration-response function of air pollution on health (Arceo, Hanna and Oliva, 2016; Pope III et al., 2011). I explore potential nonlinearities in Table A8. This table shows that alternative models fit the data no better than the linear model. If anything, the tested models suggest steeper effects at higher capacity levels.[18] I also implement a Box-Cox

[17]Column 1 shows results from a model that includes age-by-sex, month-by-year of birth (cohort), and village fixed effects. Column 2 adds plant-by-year fixed effects. The model in column 2 answers the question: when the capacity of a coal plant increases, does child height change by more in villages that are closer to the plant compared to villages that are farther away? When plant-by-year fixed effects are included, the coefficient attenuates towards zero. This is expected since it is possible that children born outside of the 50 km radius of a coal plant are also affected by the air pollution generated by the coal plant. However, the coefficient is still statistically significant. Columns 3 and 4 build off the baseline specification in Column 1 by adding birth, mother, and household characteristics. Finally, Column 5 presents a model with all control variables, and plant-by-year fixed effects.

[18]All of the models in this table build off of the specification in Table 2, column 1. Column 1 of this table simply repeats the main result from Table 2, column 1 for comparison. Column 2 allows the coefficient on capacity to be different at different quartiles of capacity, but requires the intercept to remain the same. An

power transformation on *capacity* for powers in steps of 0.1 from 0.1 to 2.0. Each power transformation is implemented in a separate regression. Figure A1 plots the resulting log-likelihoods from these models. The log likelihood is maximized just above one, indicating that effects are slightly steeper at higher capacity levels.

Table A5 tests whether the results are sensitive to dropping parts of the sample. Column 1 repeats the main analysis presented in column 1 of Table 2. In column 2, I drop the 8% of children with measured height who were born in a location other than the village in which the household was surveyed.[19] In column 3, children are included if they are born to mothers who have lived in the village in which they were interviewed for six or more years, which is true for 59% of children with measured height. In column 4, I only use expansions in existing coal plants, and in column 5, I only include children born in villages that at some point in the period are exposed to coal plants. Columns 6 and 7 drop villages in which the increase in coal plant capacity over the study period was above certain percentile thresholds. This ensures that the results are not driven by outlier villages. None of these models generate coefficients on *capacity* that are remarkably different from the effects estimated in Table 2.

6 Extensions: heterogeneity and exposure timing

6.1 Heterogeneity by socio-economic status

Is the effect of coal plant capacity on child height different for children from high SES households compared to low status households? Effects could be larger for children in poorer families because their mothers may be more likely to do work that exposes them to more pollution during pregnancy, their homes may be less insulated, their immune systems may be more compromised from other environmental risks, or their parents may be less able to seek medical attention when needed. In the context of forest fires in Indonesia, Jayachandran (2009) finds larger effects from pollution in areas with lower food consumption compared to

F-test only marginally rejects the hypothesis that these coefficients are not different from each other at the 10 percent level. Column 3 includes capacity as a quadratic. The squared term is negative, indicating that higher levels of capacity are even worse for health, but it is small in magnitude and not statistically significant. Column 4 tests whether the capacity-height relationship is characterized by diminishing marginal deficits using the natural log transformation. Because the natural log of zero is undefined, I replace ln(capacity) for unexposed children with a value of ln(0.01) so that they can be included in the regression. Column 5 uses a transformation that is defined at zero, the inverse hyperbolic sine function. In both of these regressions, coefficients on the variables of interest are at least marginally statistically significant at the 5 percent level, but they fit the data less well in the sense that they have slightly smaller adjusted R^2s compared to the model in column 1. Finally, column 6 tests a spline, which is negative indicating a steeper relationship above the median, but it is not statistically significant.

[19]These children are identifiable in the survey as children with birth dates earlier than the reported length of time the mother has lived in the village in which she was surveyed.

areas with higher food consumption.

Figure 6 shows coefficients and confidence intervals from three separate regressions interacting coal plant capacity in the month of birth with an indicator variable for whether the child meets the specified criteria for mother's literacy, mother's height, or wealth.[20] These variables were chosen because they provide an indication of household SES.[21] Notably, this figure studies heterogeneity at the household level. This is distinct from many other studies of air pollution and health that use data aggregated to the county, district, or state, in which the study of heterogeneous effects across individuals or households that have different characteristics is complicated. Regressions are of the following form:

$$\begin{aligned} height_{ivt} = &\sum_{n=1}^{N} \beta_n capacity_{vt} \times \mathbb{1}[M_{ivt} = value_n] + \\ &\sum_{n=1}^{N-1} \gamma_n \mathbb{1}[M_{ivt} = value_n] + \\ &\boldsymbol{B_{ivt}\delta} + \boldsymbol{\alpha_t} + \boldsymbol{\mu_v} + \epsilon_{ivt} \end{aligned} \tag{4}$$

where *height*, i, v, t, α and μ are as described in Equation 1. n indexes values that SES variable M can take, and N is the total number of values that M takes. The vector of birth characteristics, B, represents indicators for age-by-sex categories. Figure 6 plots coefficients β_n.

The results suggest that the effect of coal plant capacity on child height is similar among children of literate versus illiterate mothers, children born to taller versus shorter mothers, and children born into wealthy versus less wealthy families. This is in contrast to other studies which have found more severe health effects for the poor compared to the rich.

Why might both rich and poor be affected similarly in India, when they are not in other countries? One potential reason is overall poor infrastructure quality, which makes it difficult to create clean air spaces even in wealthy urban neighborhoods (Vyas, Srivastav and Spears, 2016). Another potential reason could be because high SES households tend to live closer to coal plants than low SES households, and thus are exposed to higher levels of pollution. Table 1 shows that exposed children are more likely to have literate mothers and belong to wealthier households compared to unexposed children. Figure 7 formally tests the relationship between SES and distance from a coal plant, and shows that distance from a coal plant is greater among children of illiterate mothers and children of less wealthy households.

[20] Mother's literacy and mother's height are used as controls in Section 4.3. I do not use wealth as a control in the main results because a number of characteristics that are used to construct the wealth index are included separately as controls.

[21] The DHS does not contain modules on household consumption or income.

This fact is in contrast to other contexts, where poor households are more likely to live in more polluted locations.

Since SES predicts distance from a coal plant, and the effect of coal plant capacity on child height decreases in distance from the coal plant (as seen in Figure 5), distance is an omitted interaction in the heterogeneity analysis presented in Figure 6. Figure A2 tests for heterogeneous effects by mother's literacy while also controlling for distance. I implement this test by replacing *capacity* in Equation 4 with capacity in different distance bins, from zero to 20 km, 20 to 30 km, 30 to 40 km, etc, until a distance of 70 km.[22] A clear result from this figure is that the dataset is not powered to detect differential effects by SES within distance bins. This is not surprising considering that the empirical strategy requires village fixed effects, and village alone accounts for substantial variation in literacy, mother's height, and wealth index in this sample.[23] Although the dataset is not powered to find statistically significant differences, this figure provides suggestive evidence that once distance is accounted for, the effect of coal plant capacity on child height is closer to zero among children of literate mothers compared to children of illiterate mothers.

6.2 Age of exposure

This section explores the timing of exposure, to the extent the data allow. Figure 8 shows coefficients from separate regressions of height-for-age z-score on mean capacity during various time periods, relative to birth. Each regression is of the following form:

$$height_{ivt} = \beta coalexposure_{vt} + \boldsymbol{B_{ivt}\delta} + \boldsymbol{\alpha_t} + \boldsymbol{\mu_v} + \epsilon_{ivt} \tag{5}$$

where *height*, i, v, t, α and μ are as described in Equation 1. *coalexposure* is measured in six different ways, where months are relative to birth month: capacity in the month of birth, and mean capacity in months -9 to 0, months 1 through 6, months 7 through 12, months 13 through 18, and months 19 through 24. Each of these exposure variables are tested in separate regressions.[24] The vector of birth characteristics, B, represents indicators for age-by-sex categories.

Figure 8 plots coefficients and confidence intervals. This figure provides some evidence that exposure during the 19 to 24 month period is less important for child height than exposure *in utero* or in earlier periods of life. The point estimate on exposure in the 19 to 24

[22]These are the same distance bins used in the distance analysis shown in Section 4.4.

[23]A regression with village fixed effects alone yields an $R^2 = 0.386$ when the dependent variable is literacy, $R^2 = 0.284$ for mother's height, and $R^2 = 0.646$ for the wealth index.

[24]I test these different measures of exposure in separate regressions because exposure is very highly correlated over time.

month period is positive and is not statistically different from zero. However, the standard error on this coefficient is large, and effects as large and negative as those in the month of birth cannot be rejected. In contrast, exposure to coal plants from the *in utero* period to 18 months of age are negative and statistically different from zero. Research on growth faltering among children in developing countries documents that average height-for-age z-scores decline in the first two years of life, reflecting the accumulating impact of early-life health insults on a child's growth (Victora et al., 2010). This figure provides suggestive evidence that exposure to coal plants, and the air pollution associated with them, after 18 months of age is a less stable predictor of growth faltering than exposure in earlier periods of life.

It is important to note that this figure does not necessarily prove that exposure in every month from 9 months prior to birth to 18 months after birth are relevant for child height. Capacity is highly correlated over months. For example, among children over the age of 18 months (the children comprising the sample in Figure 8) living in exposed villages, exposure in months -9 to 0 has a correlation coefficient of 0.99 with exposure in months 1 to 6, 0.98 with exposure in months 7 to 12, 0.96 with exposure in months 12 to 18, and 0.91 with exposure in months 19 to 24. These high inter-temporal correlations make it difficult to separately identify the effects of exposure in different ages. Therefore, the fact that the coefficients on exposure until 18 months are negative and statistically significant could reflect an important effect during these age ranges, or could reflect the high inter-temporal correlation in exposure.

7 Conclusion

This is the first study to demonstrate the consequences for child height of being born near a coal plant. I find that children born near a median-sized coal plant are about 0.1 standard deviations shorter than children born in the same village when there is no coal plant exposure. For an average five year old girl, this effect corresponds to a height deficit of about 0.5 cm. A battery of mechanism checks and falsification tests lends credibility to the research design. While these tests cannot directly rule out the possibility that coal capacity expansions are spuriously picking up impacts on child height through other unobserved changes, these tests suggest that the main results are not driven by other channels, and lead me to believe that they may be causal.

The effect size is small relative to the overall mean child height deficit in India of 1.48 standard deviations, compared to a healthy reference population, but it is nevertheless economically meaningful. Child height is highly correlated, at the population level, with early-life mortality, because survivors' growth is scarred by early-life disease. In the DHS, a district where children are 0.1 height-for-age standard deviations shorter would be expected, on average, to have an infant mortality that is larger by about 9 infant deaths per 1,000 live births.[25] This difference is approximately equal to two times Canada's overall infant mortality rate.

The effects documented in this study suggest that coal plants may have meaningful and enduring negative consequences for India's economy. Child height is an indicator of net nutrition early in life. Because the nutrition that contributes to linear growth also contributes to cognitive development, child height is a marker of cognitive achievement (Case and Paxson, 2008). Spears (2012) finds that Indian children who are one standard deviation taller are 5 percentage points more likely to be able to write. Applying this height-cognitive achievement association to the context studied here suggests that children exposed to an extra median-sized coal plant are about 0.5 percentage points less likely to be able to write. These results also have implications for labor market outcomes. A large literature links adult height to earnings. If the height deficits from coal plants observed here persist to adulthood, and the adult height-earnings slope remains constant, those exposed to an additional median-sized coal plant at birth would be expected to have hourly earnings that are lower by between 0.5% to 1.5%.[26]

Because coal plants are projected to continue to expand in India in the near future, the health burden that I quantify here could potentially increase unless appropriate policy action is taken to either curtail coal plant expansions, or mitigate emissions from them. Because child height has lasting consequences for human capital, the negative consequences associated with coal plants could have enduring effects for India's economy. At the very least, these negative externalities should be part of any policy debate on expanding coal plants to meet energy needs.

[25] The strong correlation between child height and early life mortality, and the literature linking air pollution with infant death (Currie et al., 2014), suggests that greater coal plant exposure may be associated with greater neonatal and infant mortality. Table A9 tests this hypothesis by regressing neonatal and infant mortality on coal plant capacity. The coefficients are in the hypothesized direction, but because early-life mortality is a relatively rare event, the estimates are noisy and imprecise. The dataset is not powered to estimate effects of coal plant capacity on infant mortality.

[26] This back-of-the-envelope calculation uses estimates of the adult earnings-height slope in various developing countries from Vogl (2014); Glick and Sahn (1998); Thomas and Strauss (1997); Schultz (2003); Haddad and Bouis (1991).

References

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

Bhati, Priyavrat, Sanjeev Kumar Kanchan, Angeline Sangeetha, Sai Siddhartha, Soundaram Ramanathan, and Abhishek Rudra. 2015. "Heat on power: Green Rating of coal-based power plants." Centre for Science and Environment.

Binkin, Nancy J, Ray Yip, Lee Fleshood, and Frederick L Trowbridge. 1988. "Birth weight and childhood growth." *Pediatrics*, 82(6): 828–834.

Bozzoli, Carlos, Angus Deaton, and Climent Quintana-Domeque. 2009. "Adult height and childhood disease." *Demography*, 46(4): 647–669.

Cameron, A Colin, and Douglas L Miller. 2015. "A practitioner's guide to cluster-robust inference." *Journal of Human Resources*, 50(2): 317–372.

Case, Anne, and Christina Paxson. 2008. "Stature and status: Height, ability, and labor market outcomes." *Journal of political Economy*, 116(3): 499–532.

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

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

Clay, Karen, Joshua Lewis, and Edson Severnini. 2016. "Canary in a coal mine: Infant mortality, property values, and tradeoffs associated with mid-20th century air pollution." National Bureau of Economic Research.

Coffey, Diane. 2015. "Prepregnancy body mass and weight gain during pregnancy in India and sub-Saharan Africa." *Proceedings of the National Academy of Sciences*, 112(11): 3302–3307.

Crimmins, Eileen M, and Caleb E Finch. 2006. "Infection, inflammation, height, and longevity." *Proceedings of the National Academy of Sciences*, 103(2): 498–503.

Currie, Janet, and Matthew Neidell. 2005. "Air pollution and infant health: what can we learn from California's recent experience?" *The Quarterly Journal of Economics*, 120(3): 1003–1030.

Currie, Janet, and Reed Walker. 2011. "Traffic congestion and infant health: Evidence from E-ZPass." *American Economic Journal: Applied Economics*, 3(1): 65–90.

Currie, Janet, and Tom Vogl. 2013. "Early-life health and adult circumstance in developing countries." *Annu. Rev. Econ.*, 5(1): 1–36.

Currie, Janet, Joshua Graff Zivin, Jamie Mullins, and Matthew Neidell. 2014. "What do we know about short-and long-term effects of early-life exposure to pollution?" *Annu. Rev. Resour. Econ.*, 6(1): 217–247.

Dey, Sagnik, Larry Di Girolamo, Aaron van Donkelaar, SN Tripathi, Tarun Gupta, and Manju Mohan. 2012. "Variability of outdoor fine particulate (PM2. 5) concentration in the Indian Subcontinent: A remote sensing approach." *Remote Sensing of Environment*, 127: 153–161.

Frankenberg, Elizabeth, Douglas McKee, and Duncan Thomas. 2005. "Health consequences of forest fires in Indonesia." *Demography*, 42(1): 109–129.

Gertler, Paul, Manisha Shah, Maria Laura Alzua, Lisa Cameron, Sebastian Martinez, and Sumeet Patil. 2015. "How does health promotion work? Evidence from the dirty business of eliminating open defecation." National Bureau of Economic Research.

Glick, Peter, and David E Sahn. 1998. "Health and productivity in a heterogeneous urban labour market." *Applied Economics*, 30(2): 203–216.

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

Haddad, Lawrence J, and Howarth E Bouis. 1991. "The impact of nutritional status on agricultural productivity: wage evidence from the Philippines." *Oxford bulletin of economics and statistics*, 53(1): 45–68.

Henderson, Vernon, Adam Storeygard, and David N Weil. 2011. "A bright idea for measuring economic growth." *American Economic Review*, 101(3): 194–99.

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

Knittel, Christopher R, Douglas L Miller, and Nicholas J Sanders. 2016. "Caution, drivers! Children present: Traffic, pollution, and infant health." *Review of Economics and Statistics*, 98(2): 350–366.

Miller, Grant, and B Piedad Urdinola. 2010. "Cyclicality, mortality, and the value of time: The case of coffee price fluctuations and child survival in Colombia." *Journal of Political Economy*, 118(1): 113–155.

Min, Brian. 2015. *Power and the vote: Elections and electricity in the developing world.* Cambridge University Press.

Pope III, C Arden, Richard T Burnett, Michelle C Turner, Aaron Cohen, Daniel Krewski, Michael Jerrett, Susan M Gapstur, and Michael J Thun. 2011. "Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure–response relationships." *Environmental health perspectives*, 119(11): 1616–1621.

Pullabhotla, Hemant. 2019. "Fires, Wind, and Smoke: Air Pollution and Infant Mortality." Working paper.

Rangel, Marcos A, and Tom S Vogl. 2018. "Agricultural Fires and Health at Birth." *Review of Economics and Statistics*, , (0).

Richards, Marcus, Rebecca Hardy, Diana Kuh, and Michael EJ Wadsworth. 2002. "Birthweight, postnatal growth and cognitive function in a national UK birth cohort." *International Journal of Epidemiology*, 31(2): 342–348.

Schultz, T Paul. 2003. "Wage rentals for reproducible human capital: evidence from Ghana and the Ivory Coast." *Economics & Human Biology*, 1(3): 331–366.

Singh, Prachi, Sagnik Dey, Sourangsu Chowdhury, and Kunal Bali. 2019. "Early Life Exposure to Outdoor Air Pollution: Effect on Child Health in India." Brookings India Working Paper 06.

Spears, Dean. 2012. "Height and cognitive achievement among Indian children." *Economics & Human Biology*, 10(2): 210–219.

Spears, Dean. 2018. "Exposure to open defecation can account for the Indian enigma of child height." *Journal of Development Economics.*

Spears, Dean, Sagnik Dey, Sourangsu Chowdhury, Noah Scovronick, Sangita Vyas, and Joshua Apte. 2019. "The association of early-life exposure to ambient PM 2.5 and later-childhood height-for-age in India: an observational study." *Environmental Health*, 18(1): 62.

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

The World Bank, Development Seed, and Brian Min. 2017. "India Night Lights."

Thomas, Duncan, and John Strauss. 1997. "Health and wages: Evidence on men and women in urban Brazil." *Journal of econometrics*, 77(1): 159–185.

van den Hooven, Edith H, Frank H Pierik, Yvonne de Kluizenaar, Albert Hofman, Sjoerd W van Ratingen, Peter YJ Zandveld, Henk Russcher, Jan Lindemans, Henk ME Miedema, Eric AP Steegers, et al. 2012. "Air pollution exposure and markers of placental growth and function: the generation R study." *Environmental health perspectives*, 120(12): 1753–1759.

Victora, Cesar Gomes, Mercedes De Onis, Pedro Curi Hallal, Monika Blössner, and Roger Shrimpton. 2010. "Worldwide timing of growth faltering: revisiting implications for interventions." *Pediatrics*, 125(3): e473.

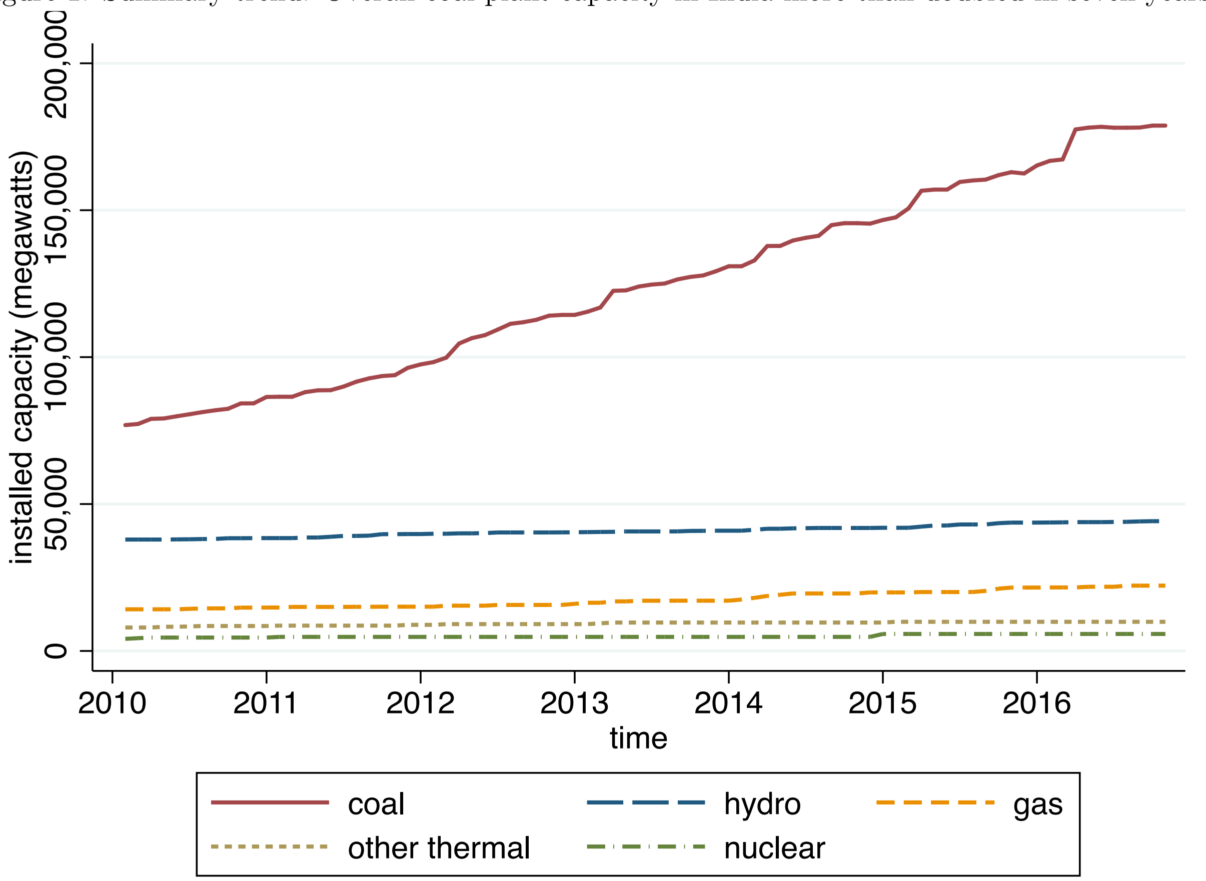
Vogl, Tom S. 2014. "Height, skills, and labor market outcomes in Mexico." *Journal of Development Economics*, 107: 84–96.

Vyas, Sangita, Nikhil Srivastav, and Dean Spears. 2016. "An Experiment with Air Purifiers in Delhi during Winter 2015-2016." *PloS one*, 11(12): e0167999.

WHO, Multicentre Growth Reference Study Group, et al. 2006. "WHO Child Growth Standards based on length/height, weight and age." *Acta paediatrica (Oslo, Norway: 1992). Supplement*, 450: 76.

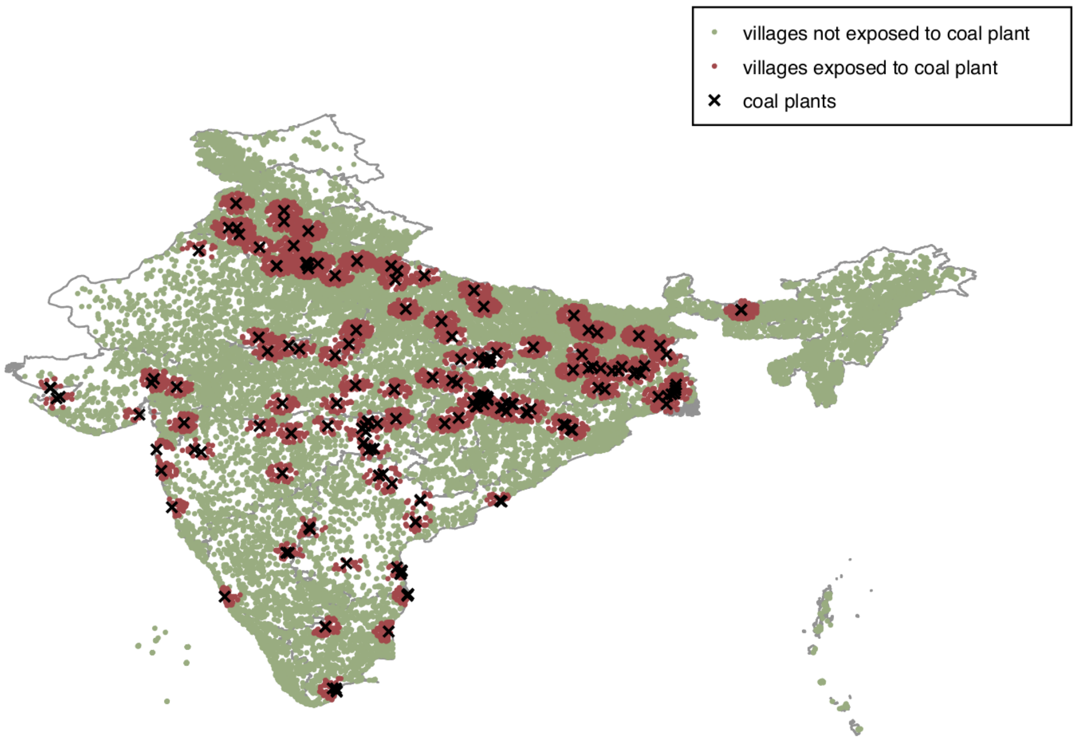
Figures

Figure 1: Summary trend: Overall coal plant capacity in India more than doubled in seven years



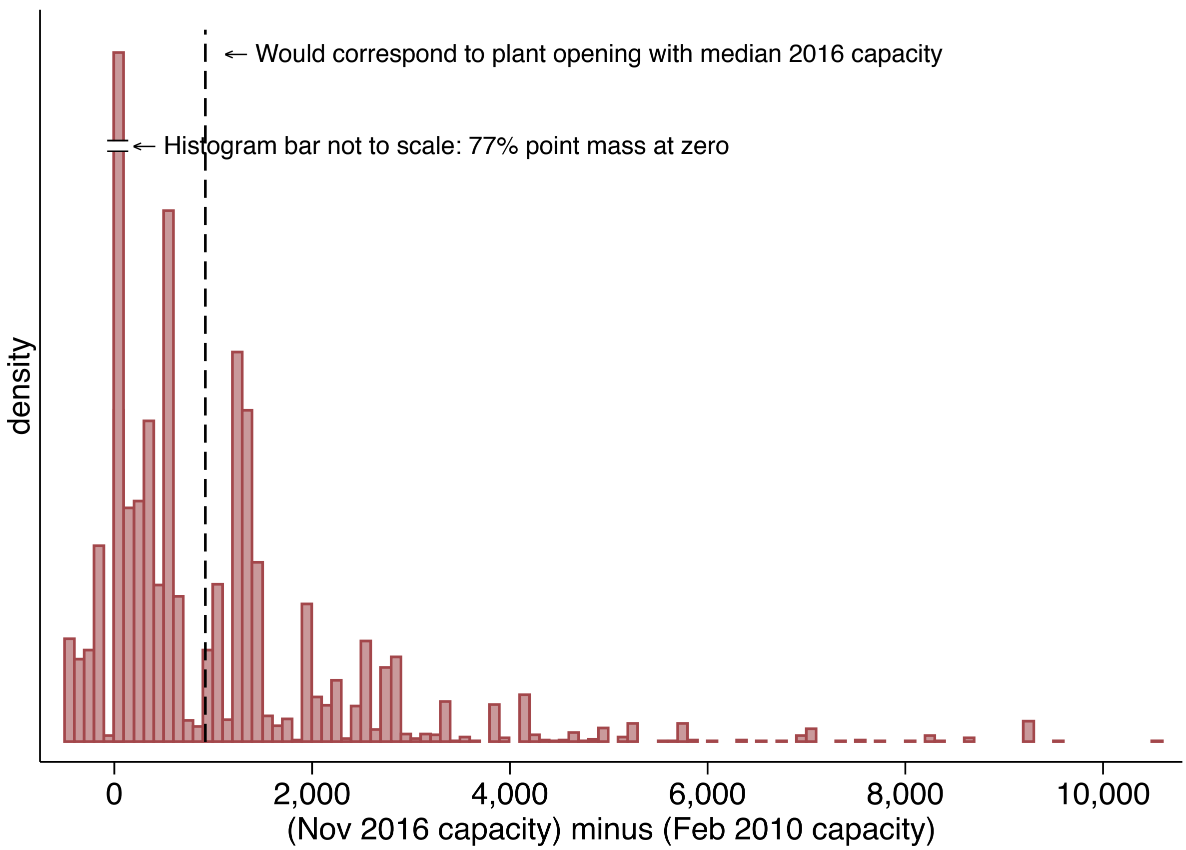
The figure displays the installed electricity-generating capacity in India from 2010 to 2016 (the period under study in this paper), by fuel source. Source: Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Figure 2: Coal plants are scattered across India



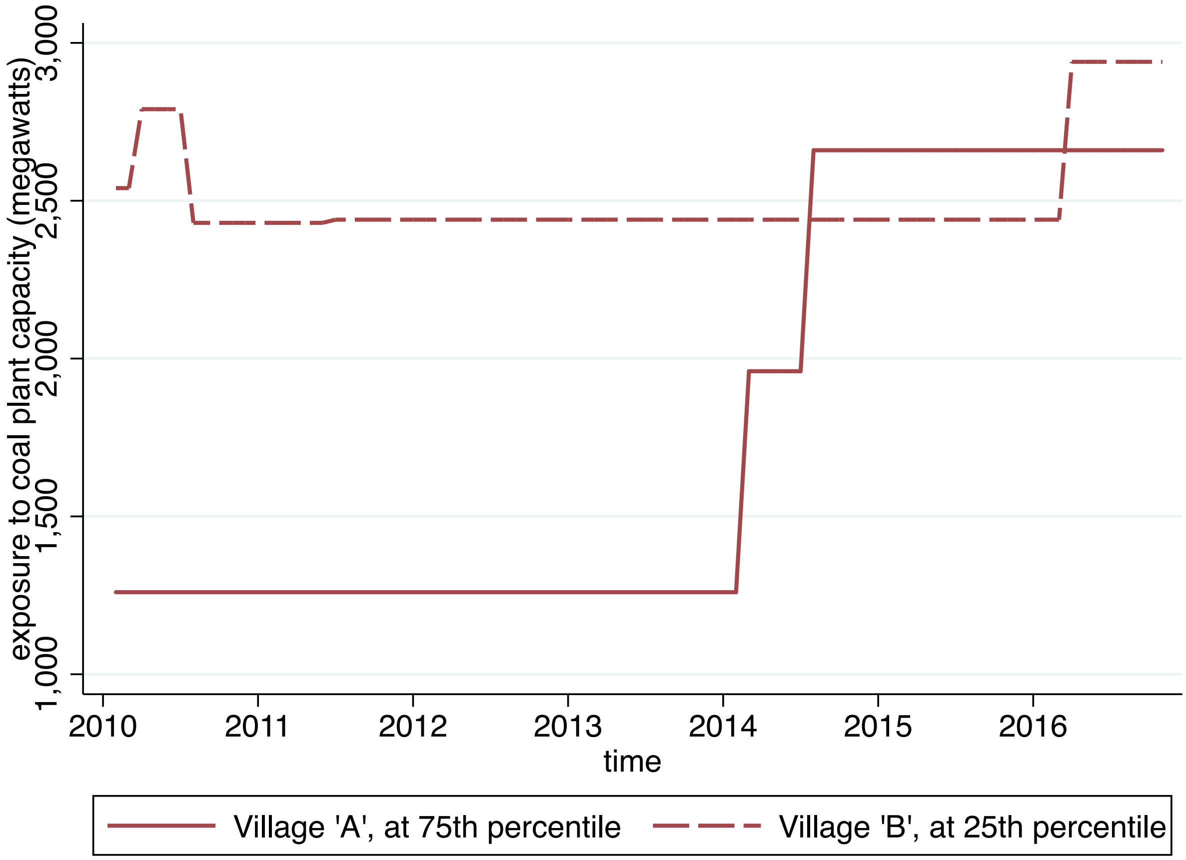
The figure maps all Indian coal plants that were installed by December 2016, and DHS urban blocks and rural villages (hereafter called villages for simplicity) by exposure status. A village is classified as exposed if it is within 50 kilometers of any coal plant installed prior to December 2016. A village is classified as unexposed if it is more than 50 kilometers from every coal plant installed prior to December 2016. This classification follows the prior literature and is validated in Figure 5. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Figure 3: Identifying variation: trends in coal plant capacity over time across all DHS villages



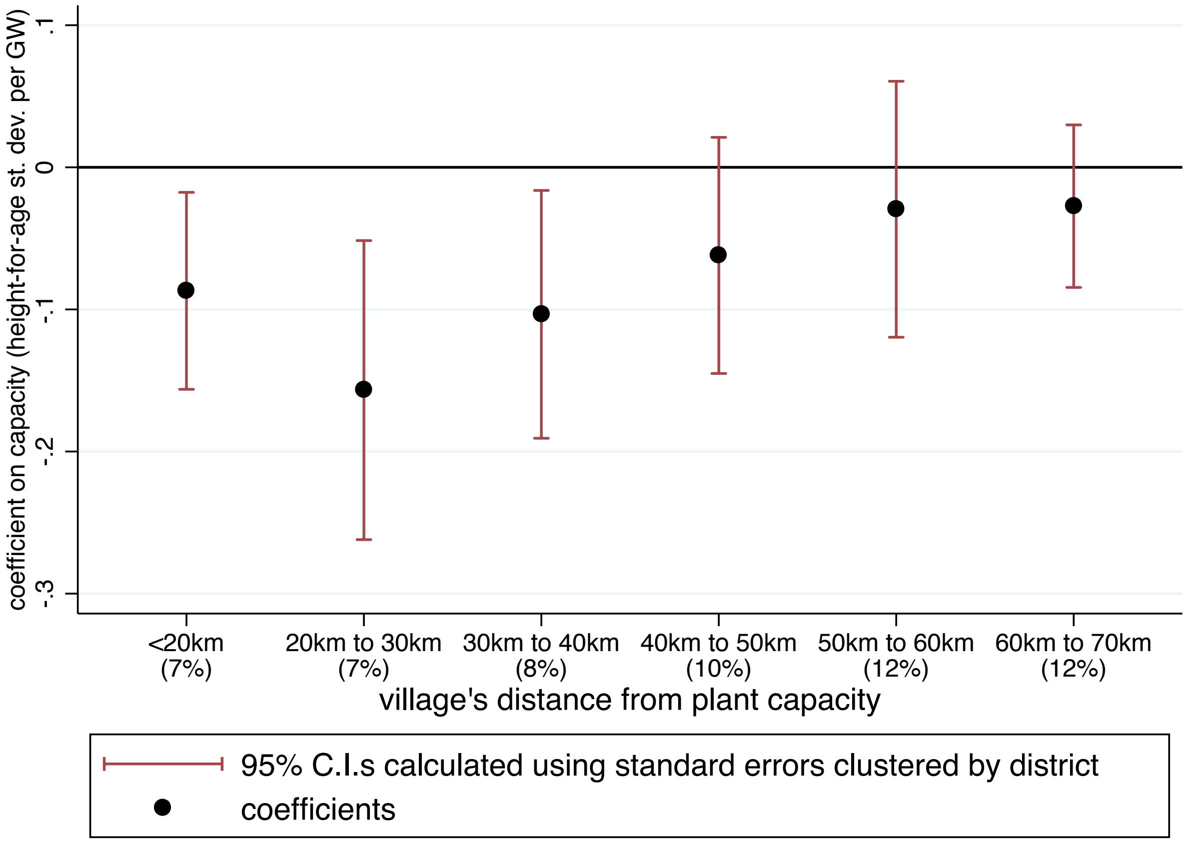
The figure displays the within-village change in coal plant capacity, where the unit of observation is the DHS village (which, for simplicity, refers to surveyed urban blocks and rural villages). Both exposed and not exposed villages are included in this figure. Because most villages were never exposed to coal plants, the histogram bar at zero is not to scale. For each village, the change is calculated by subtracting coal plant capacity exposure in February 2010 from exposure in November 2016, the period over which children with measured heights in the DHS are born. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Figure 4: Example change in coal plant exposure in two selected villages



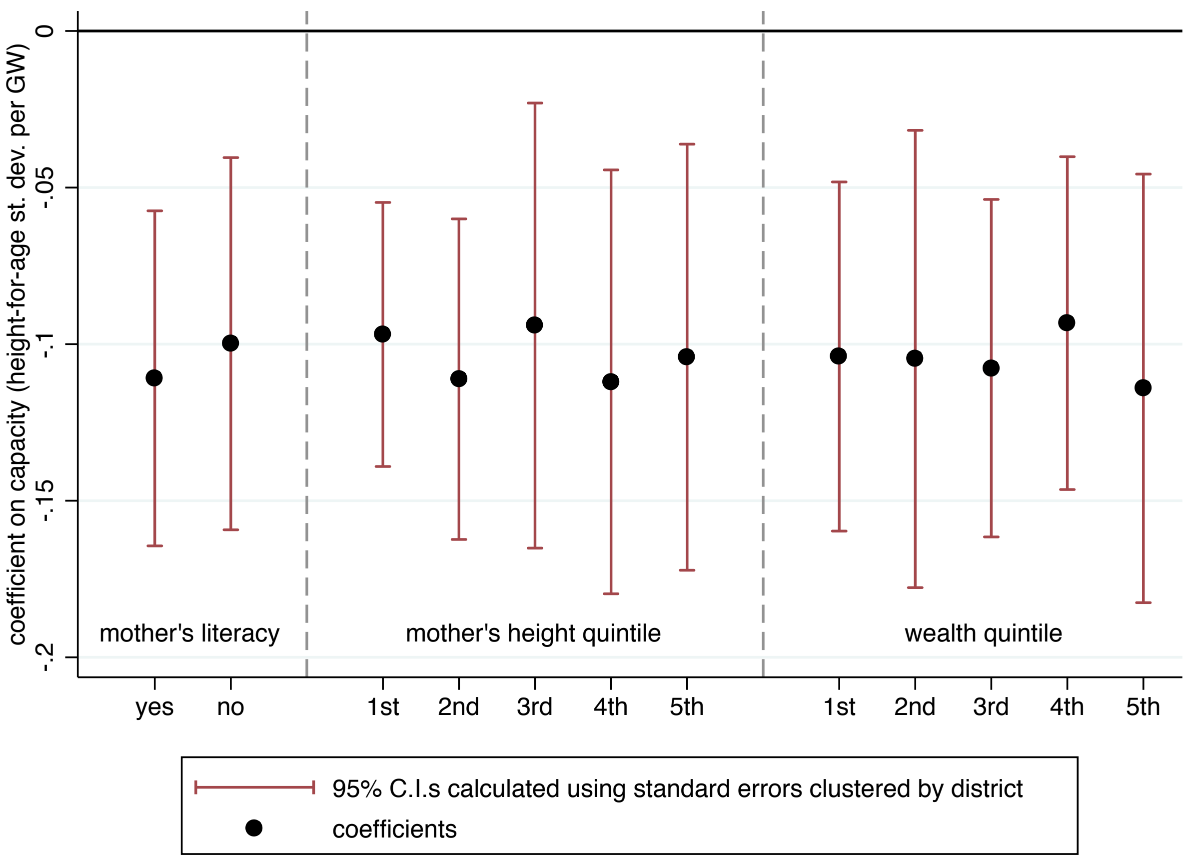
Each line shows coal plant capacity over time in a single village, one at the 25th percentile change in coal plant capacity from February 2010 to November 2016, and one at the 75th percentile change. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Figure 5: Effect of coal capacity on height attenuates with distance



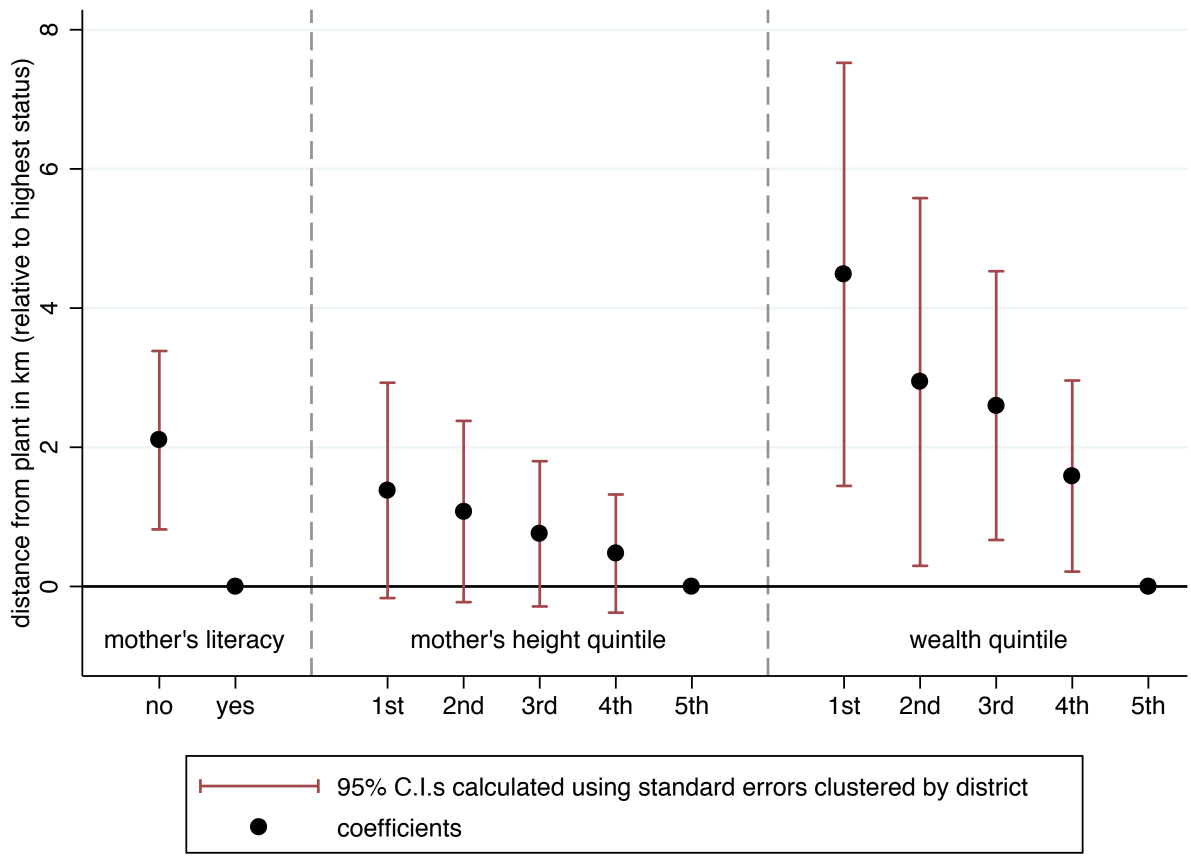
The figure displays coefficients from a single regression of height-for-age z-score on capacity in the month of birth within each of the described distance bins (see Equation 2). Sample consists of 223,166 children with measured height. Numbers in parentheses are the fraction of the sample that have positive capacity within the distance bin. Within 10 km, and between 10 and 20 km, are combined into one single bin because of small sample sizes in each of the bins separately. Bins are not mutually exclusive categories: some children are born in villages that have exposure to coal plants within multiple distance bins. Regression includes age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Figure 6: No apparent heterogeneity by socio-economic status



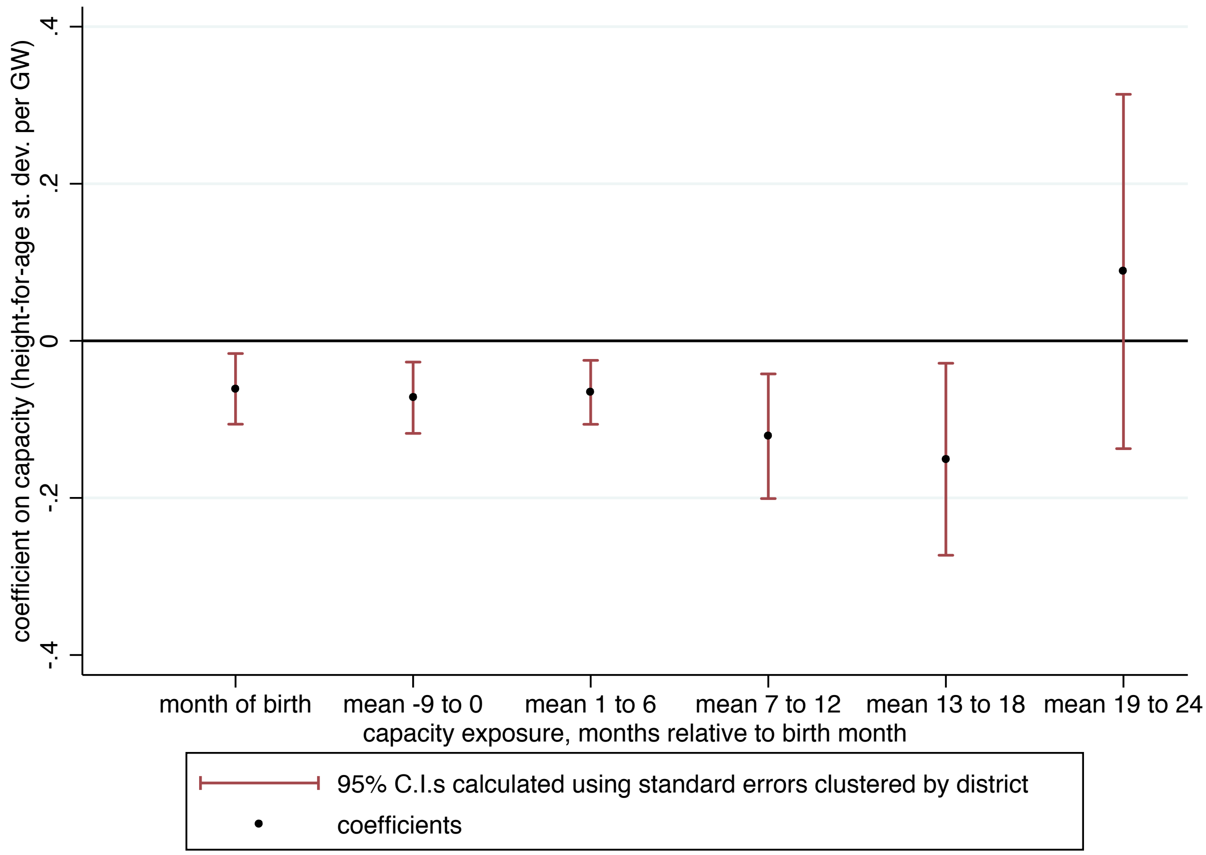
The figure displays coefficients from three separate regressions of height-for-age z-score on coal plant capacity. Each regression allows for heterogeneity in the effect of coal capacity along a different SES dimension: mother's literacy, mother's height, and household wealth index, respectively. Differential effects are estimated by interacting capacity exposure in the month of birth with indicator variables for the child meeting the specified criteria (see Equation 4). The mother's literacy sample consists of 221,575 children, the mother's height sample consists of 222,616 children, and the wealth index sample consists of 223,166 children. Sample sizes differ slightly due to data availability of heterogeneity variables. Exposed children are those in villages within 0 and 50 km of any installed coal plant. Unexposed children are those in villages farther than 50 km of all installed coal plants. Regressions include age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Figure 7: Children from lower SES households live farther from coal plants



For villages that are exposed to only one coal plant, *distance* is the distance in kilometers from the coal plant. For villages that are exposed to multiple coal plants, *distance* is a weighted average of all coal plants to which the village is exposed. Weights are the fraction of total capacity that the coal plant contributes for that village in the month of birth. The sample consists of children born in villages within 70 kilometers of any coal plant installed prior to December 2016. Regressions include age-by-sex and month-by-year of birth (cohort) fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Figure 8: The height effect of exposure to coal plant capacity after age 18 months is not statistically significant



The figure shows coefficients from separate regressions of height-for-age z-score on mean capacity during the described time periods, relative to birth. Regressions are described by Equation 5. In all regressions, the sample consists of children older than 18 months of age (n=153,855), so that exposure is defined for each category of age ranges. Exposure in the 19 to 24 month age range for children who are not yet 24 months is mean capacity for the months lived in this range. Exposed children are those in villages within 0 and 50 km of any installed coal plant. Unexposed children are those in villages farther than 50 km of all installed coal plants. The first regression uses coal capacity in the month of birth, the exposure variable used in the main analysis. The remaining regressions use mean capacity in months -9 to 0, months 1 through 6, months 7 through 12, months 13 through 18, and months 19 through 24, respectively. Effects within different age bins should be interpreted with caution since, within villages, capacity is highly correlated over months (see Figure 4). Regressions include age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Tables

Table 1: Summary Statistics

	exposure (≤50km from plant) (1)	**no exposure** (> 50km from plant) (2)	**difference** (3)	**s.e.** (4)
height-for-age	-1.489	-1.48	-0.00900	(0.0348)
capacity (GW)	1.192	0	1.192	(0.0833)
generation units	5.493	0	5.493	(0.392)
child's age (months)	30.30	29.94	0.357	(0.121)
female	0.479	0.482	-0.00241	(0.00343)
birth order	2.179	2.188	-0.00915	(0.0402)
multiple birth	0.0130	0.0145	-0.00151	(0.00103)
mom's age at birth (years)	24.25	24.32	-0.0657	(0.102)
institutional delivery	0.778	0.802	-0.0245	(0.0137)
c-section delivery	0.176	0.171	0.00462	(0.0116)
mom's height (cm)	151.4	151.8	-0.338	(0.133)
mom's literacy	0.671	0.656	0.0153	(0.0174)
Hindu	0.771	0.795	-0.0238	(0.0162)
scheduled caste	0.241	0.219	0.0224	(0.00844)
scheduled tribe	0.0774	0.123	-0.0457	(0.0101)
rural	0.643	0.762	-0.120	(0.0279)
defecates in open	0.419	0.495	-0.0763	(0.0242)
uses solid fuel	0.583	0.66	-0.0769	(0.0276)
early breastfeeding	0.672	0.695	-0.0227	(0.0150)
iron supplements in pregnancy	0.786	0.781	0.00564	(0.0142)
antihelmintics in pregnancy	0.174	0.184	-0.00977	(0.0116)
n (children under 60 months)	63,695	160,282		

The table reports child-level summary statistics for children with measured height in the DHS. Means are shown separately for children born in villages within 50 kilometers of any coal plant installed prior to December 2016, and children in villages more than 50 kilometers from every coal plant installed prior to December 2016. Capacity and units refer to coal plant exposure in the month the child was born. By construction, children born in villages with no exposure have zero capacity and units exposure in the month of birth. s.e. refers to standard errors of differences. Female, multiple birth, institutional delivery, c-section delivery, mom's literacy, Hindu, scheduled caste, scheduled tribe, rural, defecates in open, uses solid fuel, early breastfeeding, iron supplements in pregnancy, and antihelmintics in pregnancy, are binary. Means are calculated using sampling weights. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Table 2: Main result: coal plant exposure at birth is associated with height deficits

dependent variable:	**height-for-age z-score**				
	(1)	(2)	(3)	(4)	(5)
Panel A: Coal capacity (median plant capacity is 1GW)					
capacity (GW)	-0.104	-0.103	-0.0987	-0.0939	-0.105
	(0.0294)	(0.0300)	(0.0315)	(0.0328)	(0.0352)
Panel B: Coal generation units (median plant has 3 units)					
units	-0.0292	-0.0285	-0.0276	-0.0259	-0.0296
	(0.0164)	(0.0167)	(0.0169)	(0.0172)	(0.0206)
n (children under 60 months)	223,166	222,619	213,605	213,605	149,680
sex-by-age in months FE	yes	yes	yes	yes	yes
month-by-year (cohort) FE	yes	yes	yes	yes	yes
village FE	yes	yes	yes	yes	yes
birth characteristics		yes	yes	yes	yes
household characteristics			yes	yes	yes
time-by-lat and time-by-long trends				yes	
birth characteristics (last birth)					yes

The table shows fixed effects regressions described by Equation 1. Panels A and B show coefficients from two separate regressions: in panel A, the exposure variable is coal plant capacity in the month of birth, and in panel B, the exposure variable is the number of coal plant units in the month of birth. One gigawatt (GW) in coal plant capacity corresponds to the size of the median coal plant in the dataset. The median plant in the data has 3 units. The dependent variable in both panels is height-for-age z-score. Exposed children are those in villages within 0 and 50 km of any installed coal plant. Unexposed children are those in villages farther than 50 km of all installed coal plants. All specifications include sex-by-age in months, month-by-year of birth (cohort), and village fixed effects. Columns progressively add control variables. Birth characteristics include mother's age at birth, birth order, multiple birth, institutional delivery, and c-section delivery. Household characteristics include mother's height, religion, caste, literacy, household open defecation, and use of solid fuels for cooking. Birth characteristics (last birth) include whether or not the mother took iron supplements or anti-helmintics during pregnancy, variables that reduce the sample size because they were only asked of the last birth. Latitude, longitude, and time in month-years are continuous variables in the time-by-lat and time-by-long interactions. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Table 3: Falsification test: Changes in coal capacity do not predict changes in other observable child characteristics

dependent variable:	**mom's age at birth** (1)	**birth order** (2)	**multiple birth** (3)	**institutional delivery** (4)
capacity (GW)	0.0223	0.00710	-0.000792	0.00306
	(0.0434)	(0.0154)	(0.00261)	(0.00462)
n (children)	223,166	223,166	223,166	222,620
dependent variable:	**c-section** (5)	**breastfeeding w/in 1 hr of birth** (6)	**iron supplements** (7)	**drug for intestinal parasites** (8)
capacity (GW)	0.000322	-0.000924	-0.00298	0.00157
	(0.00355)	(0.00470)	(0.00515)	(0.00362)
n (children)	223,166	158,645	165,592	164,451

The table reports regressions similar to that presented in column 1 of Table 2, except that regressions presented here use as dependent variables other characteristics that are associated with child height, rather than using height-for-age z-score. Each column in this table reports the coefficient on coal plant capacity in the month of birth from a separate regression. Sample sizes differ based on the availability of the dependent variable. Exposed children are those in villages within 0 and 50 km of any installed coal plant. Unexposed children are those in villages farther than 50 km of all installed coal plants. All specifications include age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Table 4: Falsification test: Changes in coal plant capacity do not predict changes in light output at night

dependent variable:	**median monthly night lights**		
	(1)	(2)	(3)
capacity (GW)	2.850	0.457	0.164
	(0.688)	(0.294)	(0.295)
n (village-month-years)	1,007,848	1,007,848	1,007,848
st. dev. night lights	13.0	13.0	13.0
village FE		yes	yes
month-by-year (cohort) FE			yes

The table shows coefficients from regressions of median monthly light output at night on capacity. Observations are village-month-years. Data on monthly median light output at night for each Indian Census 2001 village from February 2010 through December 2013 is from The World Bank and Min (2017). Because most DHS village locations are randomly displaced by up to 5 km, night light values for DHS villages are estimated as the mean of night lights for 2001 Census villages located within 5 km of the DHS village. Column 1 reports the estimate from a univariate regression of night lights on coal plant capacity. Column 2 includes village fixed effects and column 3 includes village and month-by-year fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016, the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector, and The World Bank and Min (2017) data on light output at night for India's 2001 Census villages.

Table 5: Falsification test: Future coal plants do not predict differential height trends

dependent variable:	**height-for-age z-score**
Panel A:	
future plant × continuous time (cmc)	0.000424
	(0.00265)
continuous time (cmc)	0.0365
	(0.0208)
n (children)	159,716
Panel B:	
reference: effect of coal capacity	-0.104
(table 2, column 1)	(0.0294)

Panel A of this table reports a regression described by Equation 3. The sample in Panel A consists of children in villages that had no exposure to coal plants (villages more than 50 kilometers from every coal plant installed prior to December 2016) by the end of DHS data collection in December 2016. *future plant* is a village-level indicator for becoming exposed to a coal plant (villages within 0 and 50 km of a new coal plant) by March 2018. *continuous time* (*cmc*) is a continuous measure for month of birth, where *cmc* refers to century month code reported in the DHS. Regressions include age-by-sex, month-by-year of birth (cohort), and village fixed effects. Panel B reports the effect of coal capacity on child height from Table 2, column 1, for reference. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Appendix

Figure A1: Effects may be slightly steeper at higher capacity levels

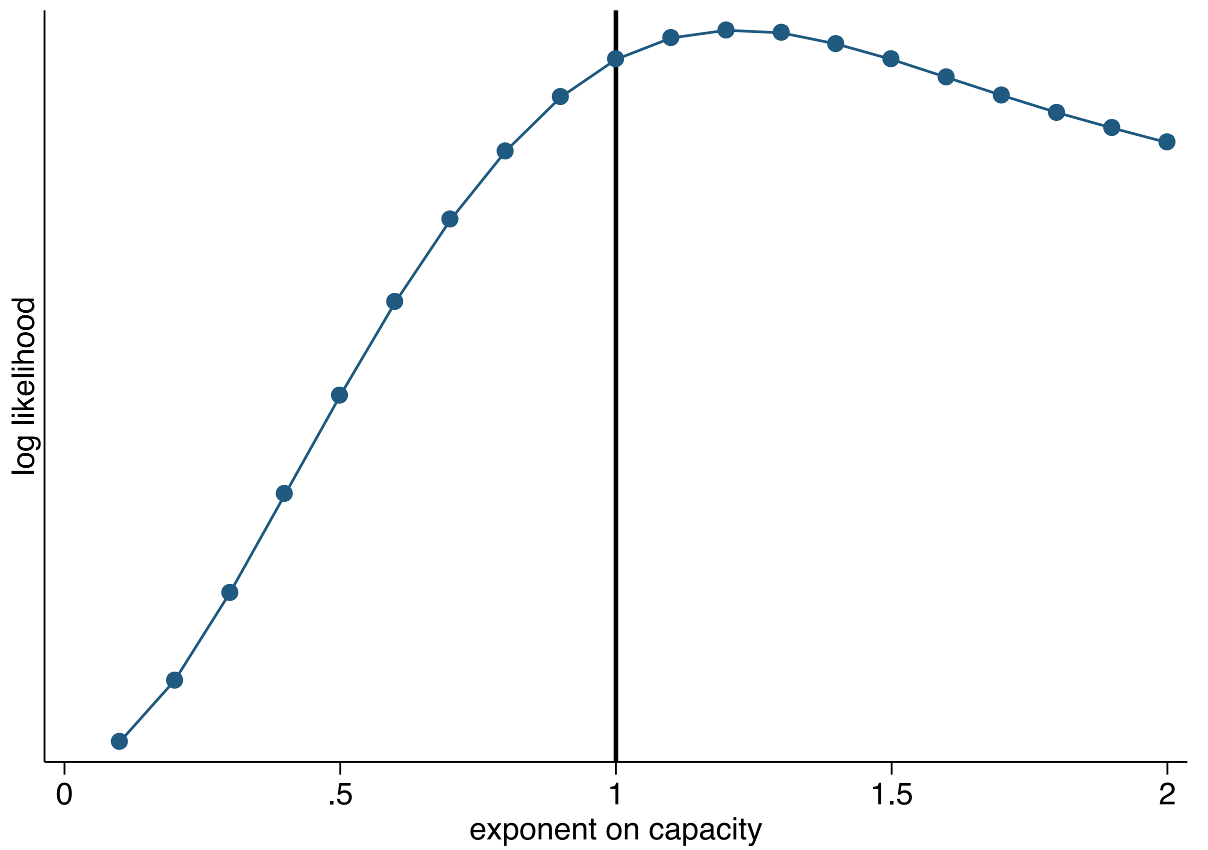
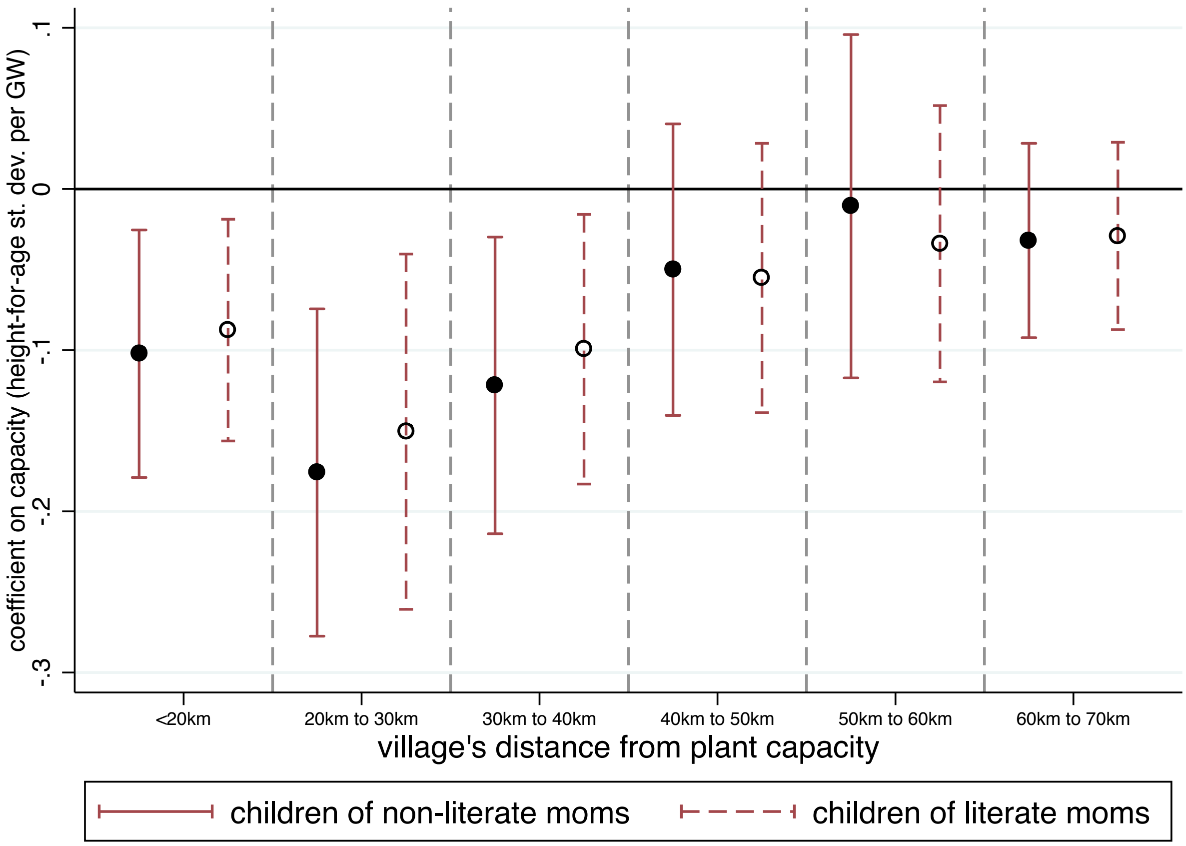


Figure displays log likelihoods for separate regressions, each with a different exponent on capacity in the month of birth, increasing in steps of 0.1 from 0.1 to 2.0. Regressions replicate the specification shown in Table 2, column 1, except that the linear form of capacity is replaced by a power transformation. Exposed children are those in villages within 0 and 50 km of any installed coal plant. Unexposed children are those in villages farther than 50 km of all installed coal plants. Each regression has a sample size of 223,166 children, and includes age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Figure A2: Dataset is not powered to detect differential effects by SES within distance bins



The figure displays coefficients from a single regression of height-for-age z-score on capacity in the month of birth within each of the described distance bins, interacted with mother's literacy. The regression is implemented by replacing *capacity* in Equation 4 with capacity in different distance bins, from zero to 20 km, 20 to 30 km, 30 to 40 km, etc, until a distance of 70 km. Distance bins are not mutually exclusive categories: some children are born in villages that have exposure to coal plants within multiple distance bins. The sample consists of 221,575 children. Regression includes age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Table A1: Coal plant capacity is associated with higher ambient air pollution within districts over time

dependent variable:	**$PM_{2.5}$ (μg/m^3)**		
	(1)	(2)	(3)
capacity (GW)	1.994	1.932	0.864
	(0.847)	(0.528)	(0.312)
n (district-month-years)	43,820	43,808	43,808
district FE		yes	yes
month-by-year (cohort) FE			yes

The table shows coefficients from regressions of $PM_{2.5}$ (μg/m^3) on capacity. Observations are district-month-years. District-level coal plant capacity for each month-year is calculated by summing capacity from all coal plants in the district, for each month-year. See Section 4.2 for more detail on the regression equation. Data for the period covering February 2010 to December 2015 are used in this analysis (pollution data after December 2015 are not available). Column 1 reports the estimate from a univariate regression of air pollution on coal plant capacity. Column 2 includes district fixed effects and column 3 includes district and month-by-year fixed effects. Sample sizes vary because some fixed effects categories lack within-category variation in the independent variable (resulting in that category being dropped). Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016, the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector, and Dey et al. (2012) data on $PM_{2.5}$ for Indian districts.

Table A2: Effect of coal capacity on height attenuates as distance increases

dependent variable:	**height-for-age z-score**
capacity (GW) × distance (km)	0.00169
	(0.000665)
capacity (GW)	-0.137
	(0.0397)
distance (km)	0.00115
	(0.000788)
n (children)	94,481

The regression results reported in this table test whether the effect of coal plant capacity in the month of birth differs by distance from the coal plant. For villages that are exposed to only one coal plant, *distance* is the distance in kilometers from the coal plant. For villages that are exposed to multiple coal plants, *distance* is a weighted average of all coal plants to which the village is exposed. Weights are the fraction of total capacity that the coal plant contributes for that village in the month of birth. *capacity* is the total coal capacity of plants within 70 km of the village. The sample consists of children born in villages within 70 kilometers of any coal plant installed prior to December 2016. Both *capacity* and *distance* are continuous variables in this regression. Regression includes age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Table A3: Associations of height-for-age with other birth and household characteristics

dependent variable:	height-for-age z-score		
	coefficient	**s.e.**	**n**
	(1)	(2)	(3)
mom's age at birth	0.00339	(0.00115)	224,188
birth order	-0.0595	(0.00373)	224,188
multiple birth	-0.238	(0.0356)	224,188
institutional delivery	0.119	(0.0132)	223,640
c-section	0.151	(0.0166)	224,188
breastfeeding w/in 1 hr of birth	-0.0243	(0.0147)	159,389
iron supplements	0.0799	(0.0163)	166,376
drug for intestinal parasites	0.0597	(0.0200)	165,230
mom is literate	0.286	(0.0132)	222,593
mom's height	0.0474	(0.00119)	223,636
wealth quintile			224,188
1st quintile	omitted	omitted	
2nd quintile	0.185	(0.0157)	
3rd quintile	0.406	(0.0207)	
4th quintile	0.587	(0.0244)	
5th quintile	0.853	(0.0297)	

The table displays coefficients from separate regressions of height-for-age z-score on various birth and household characteristics. s.e. represents standard errors. Sample sizes differ slightly due to data availability of characteristics. Regressions include age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Regressions use survey sample weights. Source: Author calculations using India's Demographic and Health Survey 2015-2016.

Table A4: Light output at night is not an omitted variable in analysis of child height and coal capacity

dependent variable:	**height-for-age z-score**		
	(1)	(2)	(3))
capacity (GW)	-0.104	-0.0727	-0.0726
	(0.0294)	(0.0392)	(0.0392)
median monthly night light			-0.000603
			(0.00102)
n (children)	223,166	106,537	106,537
sex-by-age in months FE	yes	yes	yes
month-by-year (cohort) FE	yes	yes	yes
village FE	yes	yes	yes

The table shows fixed effects regressions described by a simplified version of Equation 1. Column 1 repeats the main result from column 1 of Table 2. Column 2 shows estimates from the same regression as column 1, but limits the sample to children for which night light data is available for the village and month of birth. Column 3 adds median monthly night lights as a control variable. Data on monthly median light output at night for each Indian Census 2001 village from February 2010 through December 2013 is from The World Bank and Min (2017). Because most DHS village locations are randomly displaced by up to 5 km, night light values for DHS villages are estimated as the mean of night lights for 2001 Census villages located within 5 km of the DHS village. The dependent variable in all models is height-for-age z-score. Exposed children are those in villages within 0 and 50 km of any installed coal plant. Unexposed children are those in villages farther than 50 km of all installed coal plants. All specifications include sex-by-age in months, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016, the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector, and The World Bank and Min (2017) data on light output at night for India's 2001 Census villages.

Table A5: The main effect is not driven by particular subsamples

	dependent variable: height-for-age z-score						
sample:	**main analysis**	**born in survey village**	**mom in survey village $\geq$ 6 yrs**	**expansions in existing plants only**	**exposed villages only**	**village Δ capacity $\leq$ 99th %ile**	**village Δ capacity $\leq$ 97th %ile**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
capacity (GW)	-0.104	-0.106	-0.0959	-0.106	-0.0890	-0.106	-0.0994
	(0.0294)	(0.0304)	(0.0330)	(0.0354)	(0.0322)	(0.0252)	(0.0360)
n (children)	223,166	206,053	129,584	202,718	63,450	221,051	217,423

This table reports regressions similar to that presented in Table 2, column 1, except that different parts of the sample are dropped from the regression. Column 1 repeats the results from Table 2, column 1 for comparison. Column 2 only includes children born in the same village in which the household was interviewed by DHS surveyors. Column 3 only includes children born to mothers who have lived in the same location, where the household was interviewed, for more than five years. Column 4 only includes villages that experienced coal plant expansions, rather than new sites, as well as unexposed villages. Column 5 only includes villages within 50 kilometers of any coal plant installed prior to December 2016. Columns 6 and 7 drop villages with increases in capacity above varying thresholds. Regressions include age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Table A6: Alternative unexposed group: villages within 50 and 100 km of coal plant

dependent variable:	**height-for-age z-score**				
	(1)	(2)	(3)	(4)	(5)
Panel A: Coal capacity (median plant capacity is 1GW)					
capacity (gigwatts)	-0.101	-0.0610	-0.101	-0.0996	-0.0558
	(0.0225)	(0.0271)	(0.0227)	(0.0233)	(0.0281)
Panel B: Coal units (median plant has 3 units)					
units	-0.0263	-0.0159	-0.0256	-0.0260	-0.0136
	(0.0116)	(0.0128)	(0.0116)	(0.0117)	(0.0132)
n (children under 60 months)	132,181	132,157	131,866	128,981	128,957
sex-by-age in months FE	yes	yes	yes	yes	yes
month-by-year (cohort) FE	yes		yes	yes	
village FE	yes	yes	yes	yes	yes
plant-by-year FE		yes			yes
birth characteristics			yes	yes	yes
household characteristics				yes	yes

The table shows fixed effects regressions similar to those described by Equation 1, and is comparable to Table 2, except that the unexposed group consists of children in villages farther than 50 km of all coal plants, and within 50 and 100 kilometers of at least one coal plant installed by December 2016. Exposed children are those in villages within 0 and 50 km of any installed coal plant. Panels A and B show coefficients from two separate regressions: in panel A, the exposure variable is coal plant capacity in the month of birth, and in panel B, the exposure variable is the number of coal plant units in the month of birth. One gigawatt (GW) in coal plant capacity corresponds to the size of the median coal plant in the dataset. The median plant in the data has 3 units. The dependent variable in both panels is height-for-age z-score. All specifications include sex-by-age in months and cohort fixed effects. Column 1 is analogous to Table 2, column 1. Column 2 replaces cohort fixed effects with plant-by-year fixed effects. Columns 3 and 4 go back to the original cohort fixed effects and progressively add control variables. Column 5 includes all control variables and replaces cohort fixed effects with plant-by-year fixed effects. Birth characteristics include mother's age at birth, birth order, multiple birth, institutional delivery, and c-section delivery. Household characteristics include mother's height, religion, caste, literacy, household open defecation, and use of solid fuels for cooking. Standard errors clustered by nearest plant. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Table A7: Summary Statistics: alternative unexposed group

	exposure (≤50km from plant) (1)	**no exposure** ∈(50, 100] (2)	**difference** (3)	**s.e.** (4)
height-for-age	-1.489	-1.604	0.115	(0.0342)
capacity (GW)	1.192	0	1.192	(0.104)
generation units	5.493	0	5.493	(0.477)
child's age (months)	30.30	29.87	0.423	(0.119)
female	0.479	0.481	-0.00154	(0.00370)
birth order	2.179	2.29	-0.111	(0.0368)
multiple birth	0.0130	0.0133	-0.000387	(0.00110)
mom's age at birth (years)	24.25	24.33	-0.0810	(0.0889)
institutional delivery	0.778	0.761	0.0162	(0.0139)
c-section delivery	0.176	0.144	0.0319	(0.00925)
mom's height (cm)	151.4	151.2	0.238	(0.107)
mom's literacy	0.671	0.611	0.0605	(0.0165)
Hindu	0.771	0.808	-0.0368	(0.0161)
scheduled caste	0.241	0.231	0.00999	(0.00848)
scheduled tribe	0.0774	0.106	-0.0285	(0.0107)
rural	0.643	0.797	-0.154	(0.0377)
defecates in open	0.419	0.552	-0.134	(0.0248)
uses solid fuel	0.583	0.706	-0.122	(0.0344)
early breastfeeding	0.672	0.662	0.0103	(0.0155)
iron supplements in pregnancy	0.786	0.75	0.0368	(0.0125)
antihelmintics in pregnancy	0.174	0.17	0.00406	(0.0112)
n (children under 60 months)	63,695	132,599		

The table reports child-level summary statistics for children with measured height in the DHS. Means are shown separately for children born in villages within 50 kilometers of any coal plant installed prior to December 2016, and children in villages farther than 50 km of all coal plants, and within 50 and 100 kilometers of at least one coal plant installed by December 2016. This table is analogous to Table 1, except in how the unexposed group is defined. Capacity and units refer to coal plant exposure in the month the child was born. By construction, children born in villages with no exposure have zero capacity and units exposure in the month of birth. s.e. refers to standard errors of differences. Female, multiple birth, institutional delivery, c-section delivery, mom's literacy, Hindu, scheduled caste, scheduled tribe, rural, defecates in open, uses solid fuel, early breastfeeding, iron supplements in pregnancy, and antihelmintics in pregnancy, are binary. Means are calculated using sampling weights. Standard errors clustered by nearest plant. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Table A8: Testing linearity: Alternative models fit the data no better than the linear model

dependent variable:	**height-for-age z-score**					
	(1)	(2)	(3)	(4)	(5)	(6)
capacity (GW)	-0.104		-0.0698			-0.100
	(0.0294)		(0.0302)			(0.0562)
capacity × 1[1st quartile]		0.123				
		(0.168)				
capacity × 1[2nd quartile]		-0.108				
		(0.0797)				
capacity × 1[3rd quartile]		-0.0587				
		(0.0501)				
capacity × 1[4th quartile]		-0.0964				
		(0.0316)				
capacity^2			-0.00322			
			(0.00222)			
ln(capacity)				-0.0249		
				(0.0131)		
sinh^{-1}(capacity)					-0.163	
					(0.0484)	
above median spline						-0.00386
						(0.0688)
n (children)	223,166	224,188	223,166	224,188	223,166	223,166
F-statistic $\beta^{\text{1st q}} = \beta^{\text{2nd q}} = \beta^{\text{3rd q}} = \beta^{\text{4th q}}$		1.737				
p-value		0.158				

This table reports regressions similar to that presented in Table 2, column 1, except that the linear capacity term is replaced with different transformations of capacity. Column 1 replicates column 1 of Table 2 for reference. Column 2 allows the coefficient on capacity to be different at different quartiles of capacity, but requires the intercept to remain the same. Column 3 includes capacity as a quadratic. Column 4 tests whether the capacity-height relationship is characterized by diminishing marginal deficits using the natural log transformation. $capacity = 0.01$ replaces $capacity = 0$ in this regression because $\ln(0)$ is undefined. Column 5 uses a transformation that is defined at zero, the inverse hyperbolic sine function. Column 6 tests an above-median spline. Regressions include age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Table A9: The dataset is not powered to detect effects of coal plant capacity on neonatal or infant mortality

dependent variable:	**neonatal mortality** (1)	**infant mortality** (2)
capacity (GW)	0.0224	0.550
	(0.672)	(0.967)
n (births)	1,307,732	1,259,378
month-by-year (cohort) FE	yes	yes
village FE	yes	yes

The table reports regressions of neonatal and infant mortality on coal plant capacity in the month of birth. Sample sizes differ because children who were alive at the time of the survey are only included in the analysis if they had already exited the exposure period, which is the first month of life in column 1, and the first year of life in column 2. Exposed children are those in villages within 0 and 50 km of any installed coal plant. Unexposed children are those in villages farther than 50 km of all installed coal plants. All specifications include month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.