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

Sangita Vyas*

September 3, 2019

Abstract

This paper investigates the child health impacts associated with a large coal plant expansion in India. Using place and cohort fixed effects, exposure to a median-sized coal plant at birth is associated with a height deficit of 0.09-0.10 standard deviations. This effect is consistent with the underlying mechanism of air pollution: coal plant capacity expansions are associated with increases in air pollution, and the child height deficit associated with increases in capacity is decreasing in distance from the coal plant. The results are not driven by changes in other observable characteristics, and height pre-trends are similar in places that receive coal plants and those that do not. Heterogeneity analyses find similar effects on children of rich and poor households, but rich households are more likely to live closer to coal plants than poor households.

*University of Texas at Austin, sangita.vyas@utexas.edu. 2225 Speedway, Austin, TX 78712. This paper benefited from helpful comments received at PAA 2019. I thank Sam Arenberg, Diane Coffey, Mike Geruso, Aashish Gupta, Kevin Kuruc, Leigh Linden, Melissa Lopalo, Seth Neller, Gerald Oettinger, Dean Spears, 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 found reductions in infant mortality post policy change that are substantially larger than effect sizes found in developed countries (Chay and Greenstone, 2003; Currie and Neidell, 2005). 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.

Among developing countries, India is one of the largest consumers of coal. Over the past decade, coal plant capacity has increased dramatically, 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 present potentially large negative health externalities. This study seeks to investigate the impacts on child health that can be attributed to this dramatic expansion in coal plant capacity.

I utilize variation in the timing and geography of coal plant capacity additions in India to identify the effect of coal plant exposure on child height in a difference-in-differences framework. In particular, I link survey clusters from India's most recent Demographic and Health Survey (DHS) to coal plants based on proximity. Clusters located within 50 kilometers (km) are considered exposed to the coal plant, while those located farther than 50 km are not exposed. My strategy uses variation in exposure within clusters over time (cluster fixed effects), controlling for cohort differences that are common across all clusters (cohort fixed effects). I find that children born in clusters exposed to an additional median-sized coal plant (in terms of capacity) are 0.09 to 0.10 standard deviations shorter than children born in the same cluster with less coal plant exposure. This effect is consistent with the underlying mechanism of air pollution: coal plant capacity expansions are associated with increases in air pollution, and the child height deficit associated with increases in capacity is decreasing in distance from the coal plant. Effects are also robust to falsification tests: coal plant capacity does not similarly predict other birth characteristics related to child height, and villages near future coal plants do not have differential height trends compared to other villages that never end up near coal plants. Heterogeneity analyses find similar effects for children of high and low socio-economic status (SES) households. However, 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 covariate 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 (Rangel and Vogl, 2016) and forest fires (Jayachandran, 2009). 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) investigates 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 the two rounds of the panel. The effect of India's coal plant expansion on other indicators of child health is an open area of investigation.

As a measure of child health, I use child height, an important economic variable that predicts cognitive development (Spears, 2012), educational attainment (Case and Paxson, 2008), and adult earnings (Ibid). 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). Although I am not able to study disease in this setting, mechanisms in the literature are consistent with an effect of air pollution on child height. In particular, air pollution has been linked to low birth weight (Currie and Walker, 2011; Rangel and Vogl, 2016), and low birth weight is associated with shorter stature in childhood (Binkin et al., 1988). Additionally, 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).

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. Second, I focus on the developing country context, where 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 is likely to be different. 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. In contrast to other contexts, 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 results. Section 5 provides evidence that the effect is consistent with the mechanism of air pollution, is not driven by changes in observable characteristics, and does not appear to be confounded by differential pre-trends in height. Section 6 discusses extensions to the main analysis including heterogeneous effects and an exploration of exposure timing. Section 7 presents further robustness checks, and Section 8 concludes.

2 Data

This research uses data on child health from India's Demographic and Health Survey (DHS) 2015-2016, a dataset on power plant openings and capacity in India from 1922 to 2018, and data on $PM_{2.5}$, particulate matter smaller than 2.5 microns in diameter, estimated from satellite measurements of aerosol optical depth. 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 standard 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.

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, 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 5.2.

The final dataset consists of district-month-year estimates of $PM_{2.5}$ from 2010 to 2015. 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). District-level statistics were

[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 areas.

extracted using the shape files of district boundaries in ArcGIS. 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 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 $\text{PM}_{2.5}$ is attenuated and does not reflect the true effect size.

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.

3 Econometric framework

I use a difference-in-differences estimation strategy to identify the effect of coal plant exposure on child height. I include village 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 for that village, controlling for trends across cohorts that are common to all villages?

I estimate regressions of the following form:

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

where i indexes individual children, h indexes households, v indexes villages, and t indexes birth cohort in month-years (e.g. March 2014). *height* is the height-for-age z-score of the

[4]GPS coordinates and district are available for all coal plants. I construct coal plant capacity at the 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 $\text{PM}_{2.5}$ over time.

[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).

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 plants often have multiple units that generate electricity, and I use the number of 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, are likely to be highly correlated over time.

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.

I indirectly test the identification assumption in multiple ways in Section 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

[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).

farther away have weaker health effects. Then, I test whether coal plant capacity predicts other observable variables that are related to child height, and show that it does not. Finally, I 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.

Figure 3 shows a histogram of the identifying variation, the change in coal plant capacity within villages from February 2010 to November 2016, the period over which children in the dataset were born. 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. 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 7, 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.

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. 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 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 plant units. The median plant in my data has three units. Each column in this table corresponds to a regression with a slightly different specification. Notably, as I add more controls to the regression, the coefficient on capacity, and on units, stays relatively constant. Across all models, an additional median-sized coal plant is associated with a height deficit of 0.09-0.10 standard deviations.

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.[7] 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

[7] 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.

regressions have smaller sample sizes than the other models.

5 Mechanism and falsification tests

5.1 Air pollution

Table A1 verifies that coal plant capacity is associated with higher levels of air pollution, as measured by $PM_{2.5}$, at the district-month-year level. 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 μg/m^3 more $PM_{2.5}$.[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 $PM_{2.5}$ estimates average over spaces that likely include exposed and unexposed regions. The true effect of coal plant capacity on $PM_{2.5}$ 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.

5.2 Distance

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.[9] 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

[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.

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

zero as distance from the coal plant increases.[10] 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.[11]

5.3 Changes in observable 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. 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.

5.4 Pre-trends

My final test studies pre-trends in child height in villages that became newly exposed to coal plant capacity after the DHS completed data collection, and in 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

[10]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).

[11]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.

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 4 shows the results of this analysis. The regression equation used for this test is:

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

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 4, 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.

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 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.[12] These variables were chosen because they provide an indication of household SES.[13] 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.

[12]Mother's literacy and mother's height are used as controls in Section 4.2. 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.

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

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 if 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. Table 5 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. 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 variable in the heterogeneity analysis presented in Figure 6. Figure A1 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.[14] 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.[15] 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. Table A3 shows similar analyses accounting for distance, exploring heterogeneity by mother's height and household wealth index. Similar to literacy, effects are not statistically different within distance bins.

[14] These are the same distance bins used in the distance analysis shown in Section 5.2.

[15] 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.

6.2 Age of exposure

This section explores the timing of exposure, to the extent the data allow. Figure 7 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.[16] The vector of birth characteristics, B, represents indicators for age-by-sex categories.

Figure 7 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 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 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 7) 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

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

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 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 A4, 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 A5. 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 A4 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 A6. 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 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

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 A2 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 A7 tests whether the results are sensitive to dropping parts of the sample. In column 1, I drop children for whom reported birth dates are before the mother's reported date of moving to the location of the survey. This is true for 8 percent of children with measured height. In column 2, I only include children born in villages that at some point in the period are exposed to coal plants. Columns 3 through 6 drop observations with capacity levels that are larger than different percentiles of capacity. None of these models generate coefficients on *capacity* that are remarkably different from the effects estimated in Table 2.

8 Conclusion

To my knowledge, this is the first study to investigate the implications for child height of being born near a coal plant. I find that children born near an extra median-sized coal plant are 0.09 to 0.10 standard deviations shorter than children born in the same village when there is less coal plant capacity. This association is robust to a number of mechanism checks and falsification tests that lend 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. This effect represents one-sixth of the height gap between children of illiterate versus literate mothers, and two-thirds of the much debated height gap between children in India and children in sub-Saharan Africa. Moreover, 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 0.10 height-for-age standard deviations shorter would be expected, on average to have an infant mortality that is larger by 8 infant deaths per 1,000 live births. This difference is approximately equal to two times the United Kingdom's overall infant mortality rate.

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

the median, but it is not statistically significant.

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.

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.

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.

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.

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

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.

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.

Rangel, Marcos A, and Tom Vogl. 2016. "Agricultural fires and infant health." National Bureau of Economic Research.

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.*

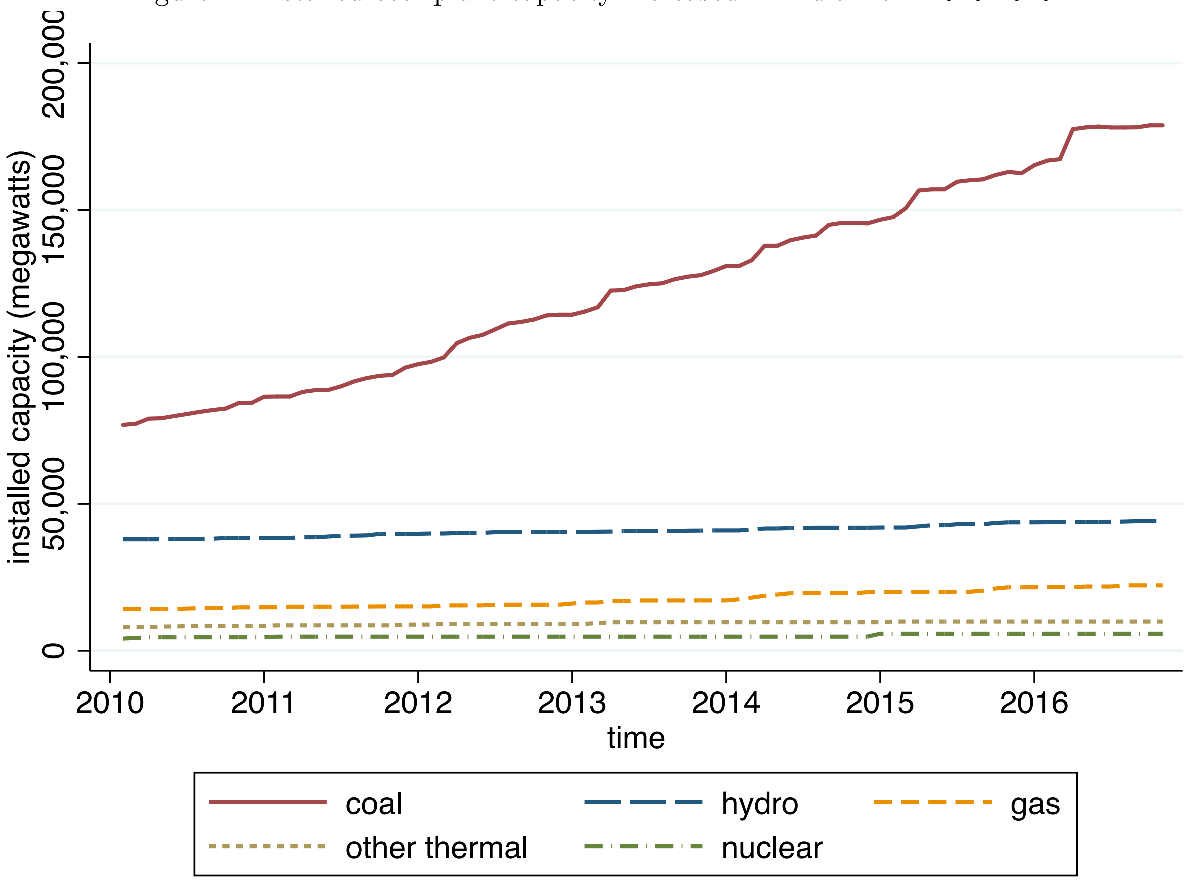
Tanaka, Shinsuke. 2015. "Environmental regulations on air pollution in China and their impact on infant mortality." *Journal of health economics*, 42: 90–103.

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.

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

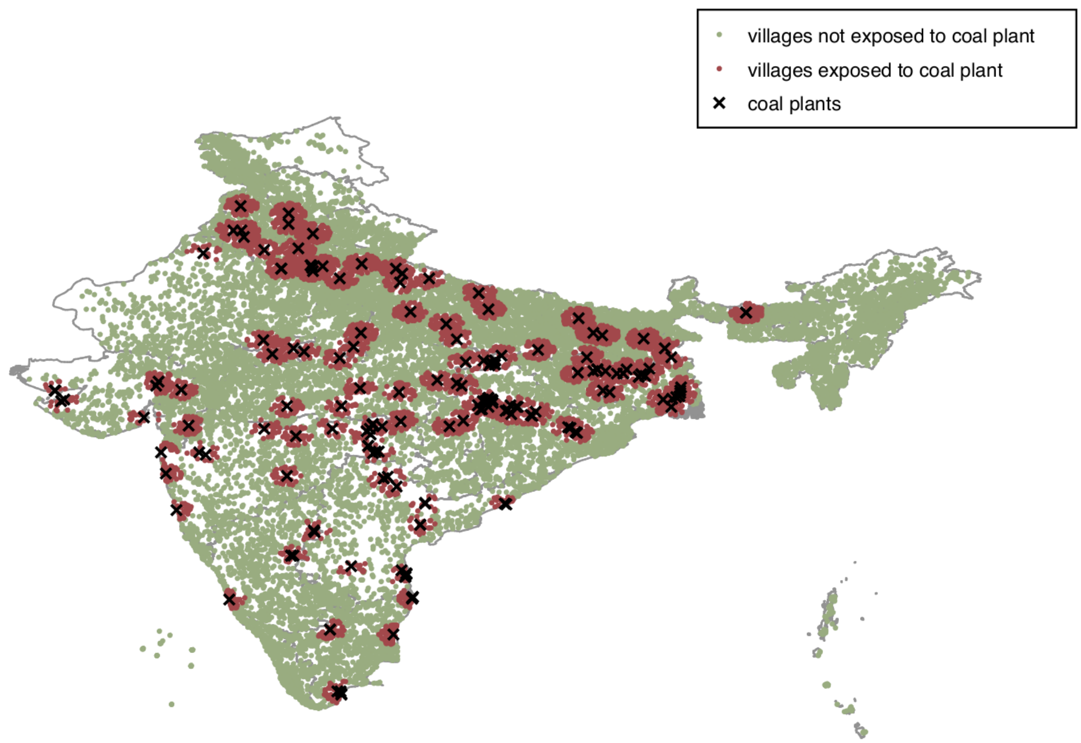
Figures

Figure 1: Installed coal plant capacity increased in India from 2010-2016



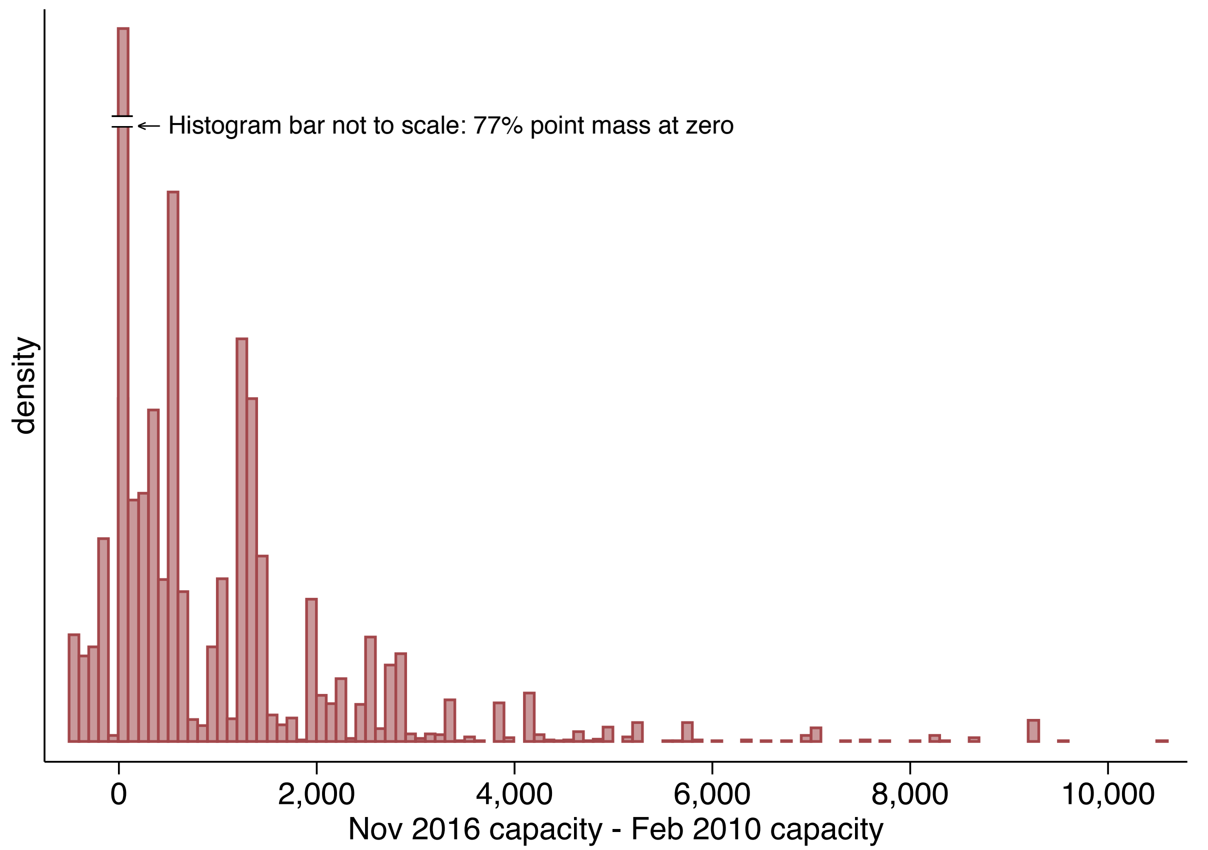
The figure displays the installed electricity-generating capacity in India from 2010 to 2016 (the period under study in this paper), by fuel source. Source: Central Electricity Authority of India's CO_2 Baseline Database for the Indian Power Sector.

Figure 2: Geography of coal plants and DHS villages



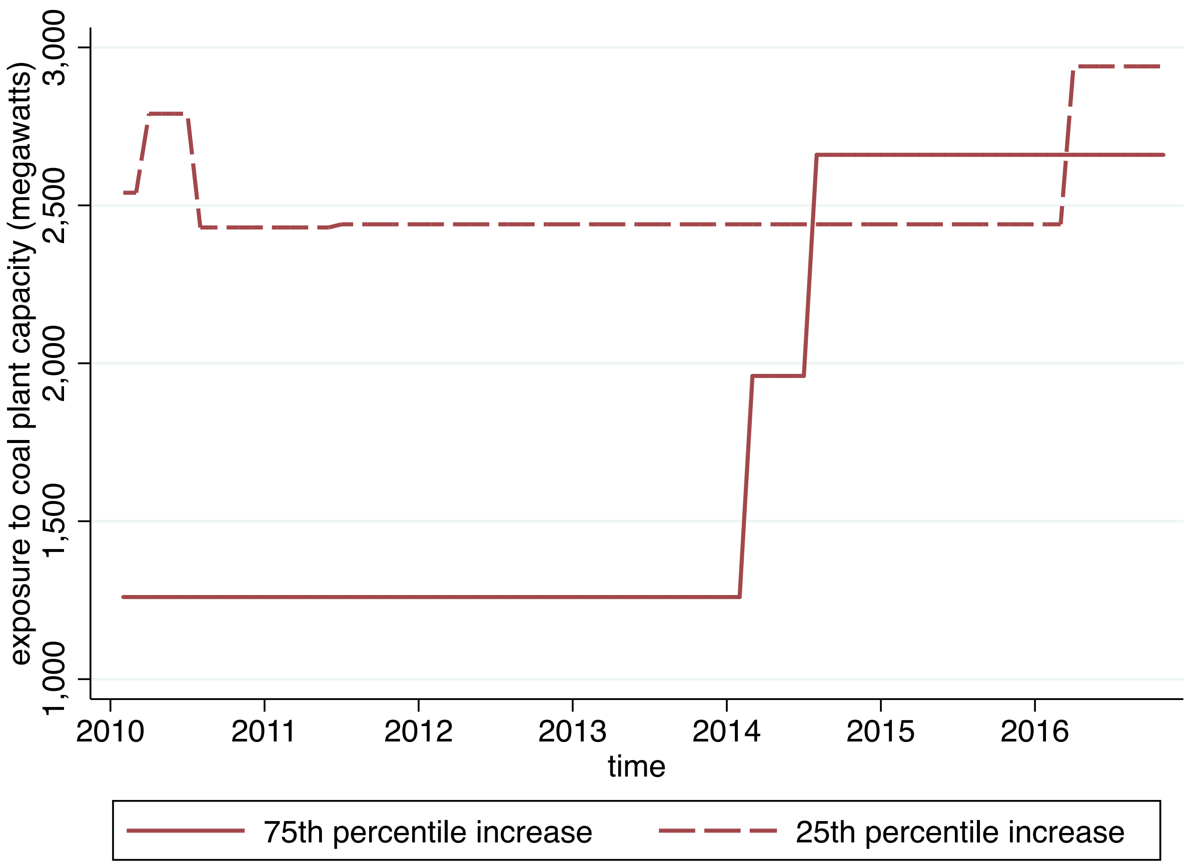
The figure maps coal plants installed by December 2016, and villages by exposure status. A village is classified as exposed if it is within 50 kilometers of any coal plant installed prior to December 2016. A village is classified as unexposed if it is more than 50 kilometers from every coal plant installed 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



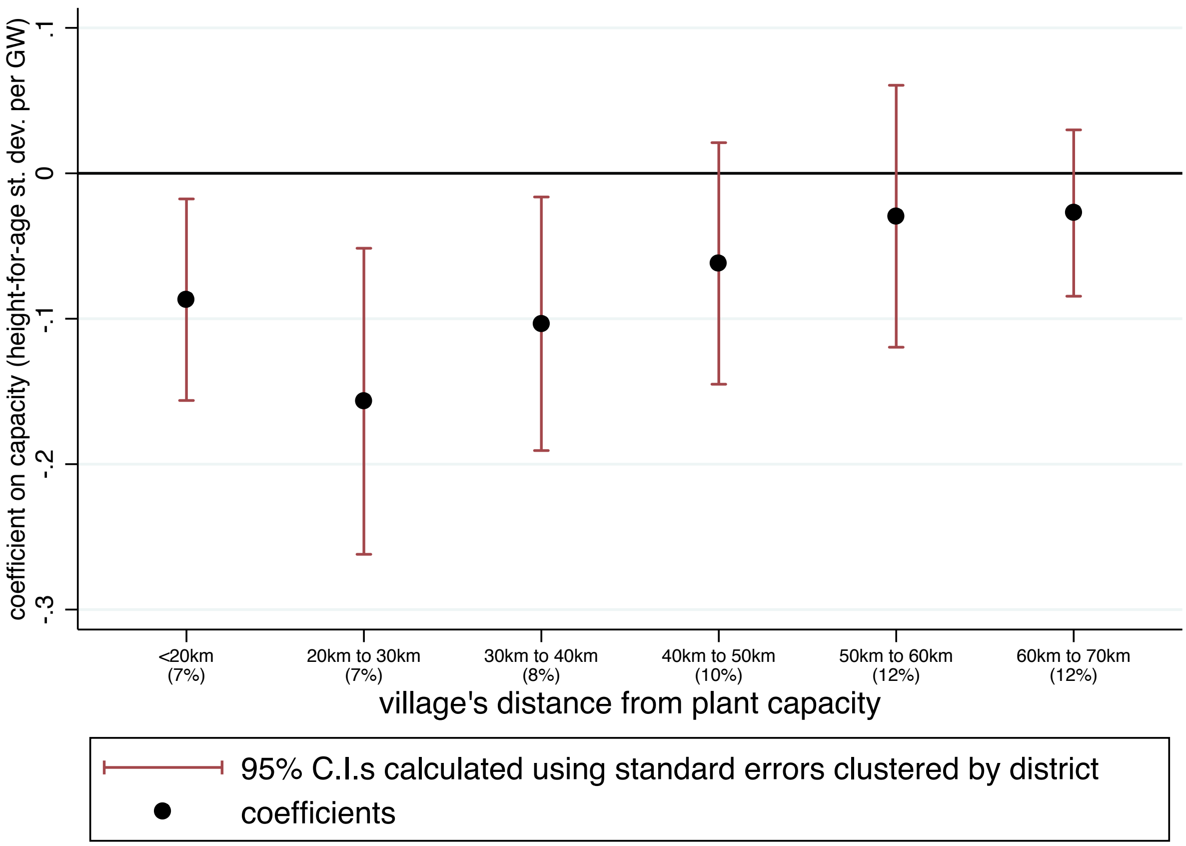
The figure displays the within-village change in coal plant capacity, where the unit of observation is the DHS villages. Both exposed and not exposed villages are included in this figure. Because most villages were never exposed to coal plants, the histogram bar at zero is not to scale. For each village, 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: Change in coal plant exposure in select villages



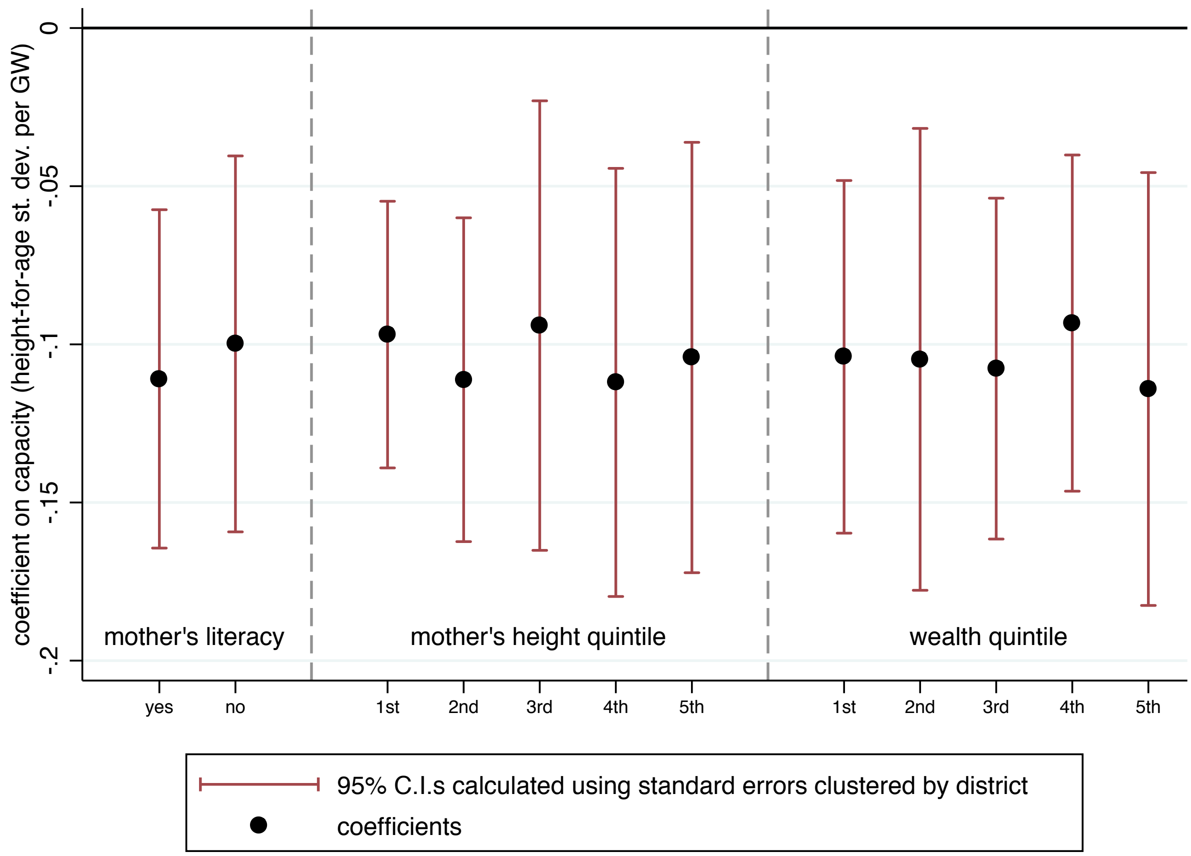
Each line shows coal plant capacity over time in a single village, one at the 25th percentile change in coal plant capacity from February 2010 to November 2016, and one at the 75th percentile change. Source: Author calculations using India's Demographic and Health Survey 2015-2016 and the Central Electricity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

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



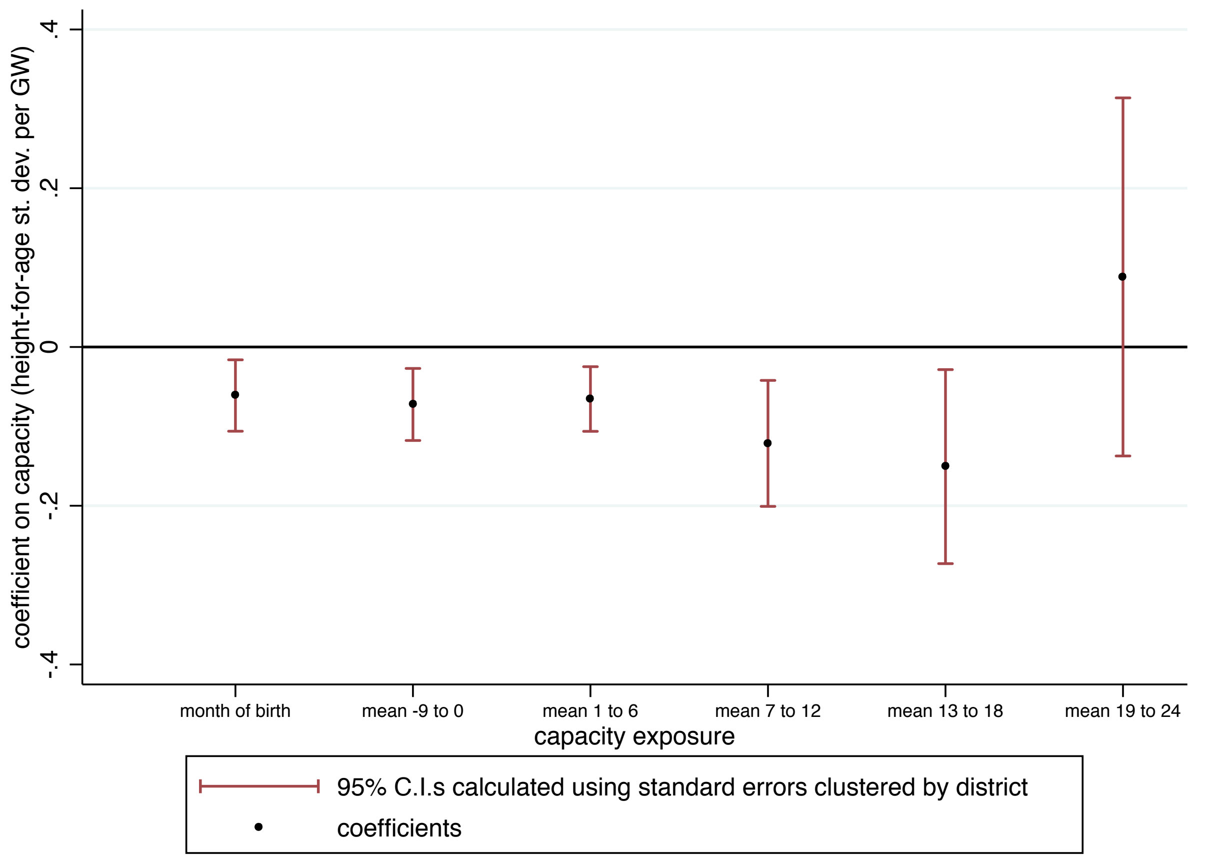
The figure displays coefficients from a single regression of height-for-age z-score on capacity in the month of birth within each of the described distance bins (see Equation 2). Sample consists of 223,166 children with measured height. Numbers in parentheses are the fraction of the sample that have 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.

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.

Figure 7: The height effect of exposure to coal plant capacity after 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.

Tables

Table 1: Summary Statistics

	exposure (≤50km from coal plant) (1)	**no exposure** (> 50km from coal plant) (2)	**difference** (3)
height-for-age	-1.489	-1.480	-0.00900
capacity (GW)	1.192	0	1.192**
units	5.493	0	5.493**
child's age (months)	30.30	29.94	0.357**
female	0.479	0.482	-0.00241
birth order	2.179	2.188	-0.00915
multiple birth	0.0130	0.0145	-0.00151
mom's age at birth (years)	24.25	24.32	-0.0657
institutional delivery	0.778	0.802	-0.0245**
c-section delivery	0.176	0.171	0.00462
mom's height (cm)	151.4	151.8	-0.338**
mom's literacy	0.671	0.656	0.0153**
Hindu	0.771	0.795	-0.0238**
scheduled caste	0.241	0.219	0.0224**
scheduled tribe	0.0774	0.123	-0.0457**
rural	0.643	0.762	-0.120**
defecates in open	0.419	0.495	-0.0763**
uses solid fuel	0.583	0.66	-0.0769**
early breastfeeding	0.672	0.695	-0.0227**
iron supplements in pregnancy	0.786	0.781	0.00564
antihelmintics in pregnancy	0.174	0.184	-0.00977*
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. 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. ** $p<0.01$, * $p<0.05$, + $p<0.10$.

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					
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 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. ** $p<0.01$, * $p<0.05$, + $p<0.10$.

Table 3: Falsification test: Coal capacity does not predict other observables

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	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	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. ** $p<0.01$, * $p<0.05$, + $p<0.10$.

Table 4: Falsification test: Future coal plants do not predict differential height trends

dependent variable:	**height-for-age z-score**
Panel A:	
future plant X continuous time (cmc)	0.000424
	(0.00265)
continuous time (cmc)	0.0365+
	(0.0208)
n	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. $futureplant$ 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. ** p<0.01, * p<0.05, + p<0.10.

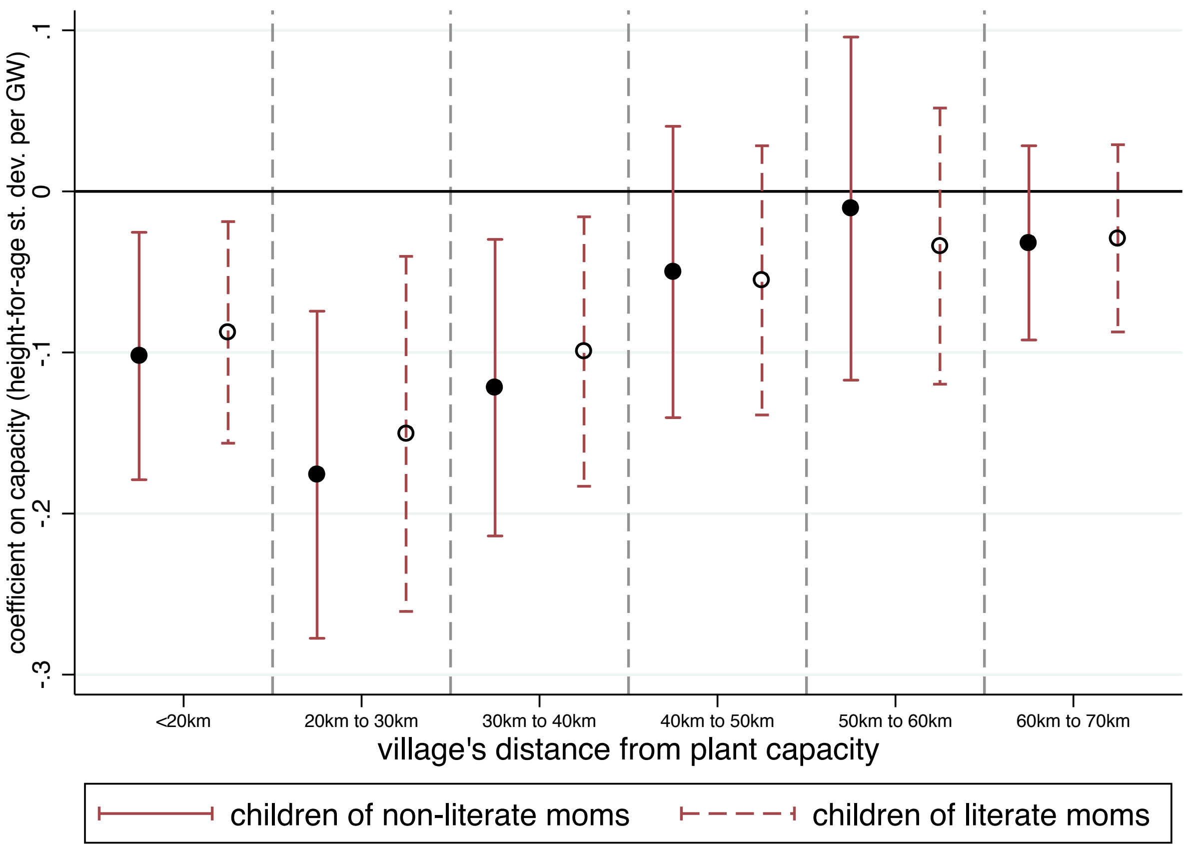
Table 5: Children from higher SES households live closer to coal plants

dependent variable:	**distance from coal plant**		
	(1)	(2)	(3)
mom not literate	2.101**		
	(0.654)		
mom literate	omitted category		
1st quintile mom's height		1.380+	
		(0.789)	
2nd quintile mom's height		1.076	
		(0.664)	
3rd quintile mom's height		0.756	
		(0.532)	
4th quintile mom's height		0.473	
		(0.433)	
5th quintile mom's height		omitted category	
1st quintile wealth index			4.484**
			(1.551)
2nd quintile wealth index			2.938*
			(1.348)
3rd quintile wealth index			2.599**
			(0.985)
4th quintile wealth index			1.586*
			(0.701)
5th quintile wealth index			omitted category
F-test (all indicators=0)		0.795	3.274
p-value		0.529	0.0117
n	94,101	94,534	94,793

The table displays coefficients of regressions of distance from a coal plant on different indicators of socio-economic status. The table explores three different measures of socio-economic status: mother's literacy in column 1, mother's height in column 2, and household wealth index in column 3. The sample consists of children born in villages within 70 kilometers of any coal plant installed prior to December 2016. 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 omitted category for the mom's literacy regression is mom is literate. The omitted categories for regressions with mom's height and household wealth index are the 5th quintiles of each, respectively. Regression includes age-by-sex and month-by-year of birth (cohort) fixed effects. Standard errors clustered by district. ** $p<0.01$, * $p<0.05$, + $p<0.10$.

Appendix

Figure A1: 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.

Figure A2: Effects are potentially steeper at higher capacity levels

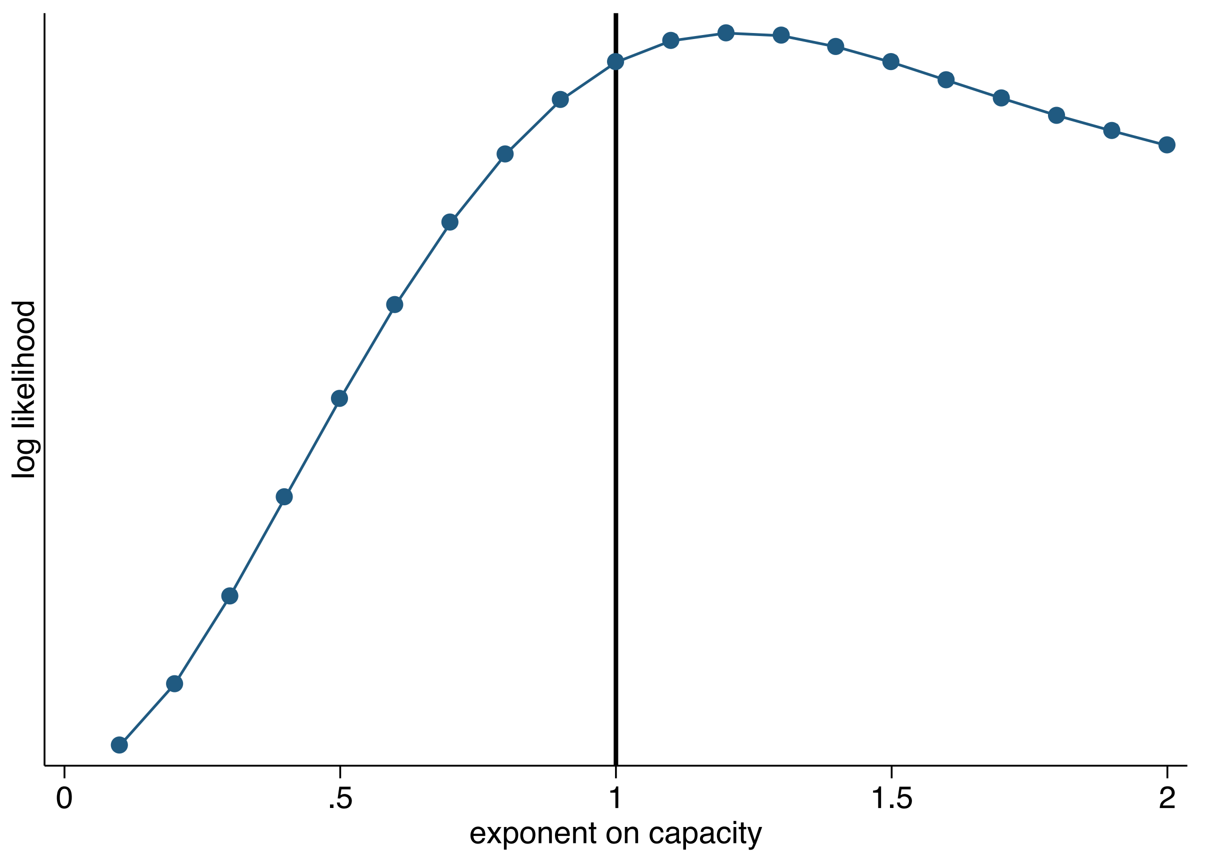


Figure displays log likelihoods for separate regressions, each with a different exponent on capacity in the month of birth, increasing in steps of 0.1 from 0.1 to 2.0. Regressions replicate the specification shown in Table 2, column 1, except that the linear form of capacity is replaced by a power transformation. 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.

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 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 5.1 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. ** p<0.01, * p<0.05, + p<0.10.

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

dependent variable:	**height-for-age z-score**
capacity (GW) X distance (km)	0.00169*
	(0.000665)
capacity (GW)	-0.137**
	(0.0397)
distance (km)	0.00115
	(0.000788)
n	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. ** $p<0.01$, * $p<0.05$, + $p<0.10$.

Table A3: Heterogeneity by distance

dependent variable:	**height-for-age z-score**	
interactions with:	**mom's height quintile**	**household wealth quintile**
	(1)	(2)
1st quintile	-0.873**	-0.851**
	(0.0157)	(0.0223)
2nd quintile	-0.578**	-0.657**
	(0.0140)	(0.0192)
3rd quintile	-0.414**	-0.447**
	(0.0140)	(0.0180)
4th quintile	-0.253**	-0.269**
	(0.0135)	(0.0179)
1st quintile X capacity $\in$ (0, 20)	-0.0591	-0.0504
	(0.0418)	(0.0469)
2nd quintile X capacity $\in$ (0, 20)	-0.0971**	-0.0561
	(0.0353)	(0.0443)
3rd quintile X capacity $\in$ (0, 20)	-0.107**	-0.112**
	(0.0362)	(0.0386)
4th quintile X capacity $\in$ (0, 20)	-0.0870*	-0.108**
	(0.0349)	(0.0392)
5th quintile X capacity $\in$ (0, 20)	-0.0645+	-0.0977**
	(0.0385)	(0.0375)
1st quintile X capacity $\in$ [20, 30)	-0.166**	-0.177**
	(0.0535)	(0.0576)
2nd quintile X capacity $\in$ [20, 30)	-0.131*	-0.213**
	(0.0513)	(0.0580)
3rd quintile X capacity $\in$ [20, 30)	-0.146*	-0.142*
	(0.0682)	(0.0599)
4th quintil X capacity $\in$ [20, 30)	-0.186**	-0.110*
	(0.0600)	(0.0547)
5th quintile X capacity $\in$ [20, 30)	-0.180**	-0.175*
	(0.0566)	(0.0685)
1st quintile X capacity $\in$ [30, 40)	-0.103**	-0.125*
	(0.0381)	(0.0502)
2nd quintile X capacity $\in$ [30, 40)	-0.112**	-0.0878
	(0.0426)	(0.0544)
3rd quintile X capacity $\in$ [30, 40)	-0.0766	-0.0854+
	(0.0524)	(0.0448)
4th quintile X capacity $\in$ [30, 40)	-0.108+	-0.130*
	(0.0561)	(0.0505)
5th quintile X capacity $\in$ [30, 40)	-0.119+	-0.123*
	(0.0609)	(0.0552)
1st quintile X capacity $\in$ [40, 50)	-0.0767+	-0.0705
	(0.0443)	(0.0515)
2nd quintile X capacity $\in$ [40, 50)	-0.105*	-0.0856+
	(0.0449)	(0.0494)
3rd quintile X capacity $\in$ [40, 50)	-0.0544	-0.0846+
	(0.0529)	(0.0468)
4th quintile X capacity $\in$ [40, 50)	-0.0804+	-0.00624
	(0.0483)	(0.0508)
5th quintile X capacity $\in$ [40, 50)	-0.0705	-0.0698
	(0.0470)	(0.0472)
n	222,616	223,166

The table displays coefficients from two separate regressions of height-for-age z-score on capacity in the month of birth within different distance bins, interacted with a variable describing socio-economic status. The regression in column 1 explores heterogeneity by mother's height, and column 2 explores heterogeneity by household wealth index. Regressions are 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 50 km. Distance bins are not mutually exclusive categories: some children are born in villages that have exposure to coal plants within multiple distance bins. Sample sizes differ slightly due to data availability of heterogeneity variables. Regressions include age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. ** p<0.01, * p<0.05, + p<0.10.

Table A4: 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					
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					
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. ** p<0.01, * p<0.05, + p<0.10.

Table A5: Summary Statistics: alternative unexposed group

	exposure (≤50km from coal plant) (1)	**no exposure** ∈(50, 100] (2)	**difference** (3)
height-for-age	-1.489	-1.604	0.115**
capacity (GW)	1.192	0	1.192**
units	5.493	0	5.493**
child's age (months)	30.30	29.87	0.423**
female	0.479	0.481	-0.00154
birth order	2.179	2.29	-0.111**
multiple birth	0.0130	0.0133	-0.000387
mom's age at birth (years)	24.25	24.33	-0.0810
institutional delivery	0.778	0.761	0.0162
c-section delivery	0.176	0.144	0.0319**
mom's height (cm)	151.4	151.2	0.238*
mom's literacy	0.671	0.611	0.0605**
Hindu	0.771	0.808	-0.0368*
scheduled caste	0.241	0.231	0.00999
scheduled tribe	0.0774	0.106	-0.0285**
rural	0.643	0.797	-0.154**
defecates in open	0.419	0.552	-0.134**
uses solid fuel	0.583	0.706	-0.122**
early breastfeeding	0.672	0.662	0.0103
iron supplements in pregnancy	0.786	0.75	0.0368**
antihelmintics in pregnancy	0.174	0.17	0.00406
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. 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. ** $p<0.01$, * $p<0.05$, + $p<0.10$.

Table A6: 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 x 1[1st quartile]		0.123				
		(0.168)				
capacity x 1[2nd quartile]		-0.108				
		(0.0797)				
capacity x 1[3rd quartile]		-0.0587				
		(0.0501)				
capacity x 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	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. ** p<0.01, * p<0.05, + p<0.10.

Table A7: The main effect is not driven by outliers

sample:	**dependent variable: height-for-age z-score**					
	born in place of interview	**exposed villages only**	**children exposed to capacity $\leq$**			
			99th %ile	**95th %ile**	**90th %ile**	**75th %ile**
	(1)	(2)	(3)	(4)	(5)	(6)
capacity (GW)	-0.106**	-0.0890**	-0.0927**	-0.102**	-0.0677+	-0.0911+
	(0.0304)	(0.0322)	(0.0262)	(0.0277)	(0.0350)	(0.0482)
n	206,053	63,450	222,157	220,234	218,383	209,578

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 only includes children born in the same village in which the household was interviewed by the DHS. Column 2 only includes villages within 50 kilometers of any coal plant installed prior to December 2016. Columns 3 through 6 drop observations with extreme capacity values above different thresholds. Regressions include age-by-sex, month-by-year of birth (cohort), and village fixed effects. Standard errors clustered by district. ** p<0.01, * p<0.05, + p<0.10.