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

Sangita Vyas*

July 11, 2019

Abstract

I investigate the child health impacts associated with a large coal plant expansion in India. Using place and time fixed effects, exposure to a median-sized coal plant at birth is associated with a height deficit of 0.09 standard deviations. Effects are similar for low and high socio-economic status children. I show evidence supporting air pollution as the mechanism underlying this relationship. I also demonstrate that the results are not driven by changes in other characteristics, and that height pre-trends are similar in places that receive coal plants and those that do not. Because child height has lasting consequences for human capital, coal plants likely have enduring effects for India's economy.

*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.

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).

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 generalized 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 are considered exposed to the coal plant, while those located farther than 50 kilometers are not exposed. My strategy uses variation in exposure within clusters over time (cluster fixed effects), controlling for secular time trends common across all clusters (time 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 village 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

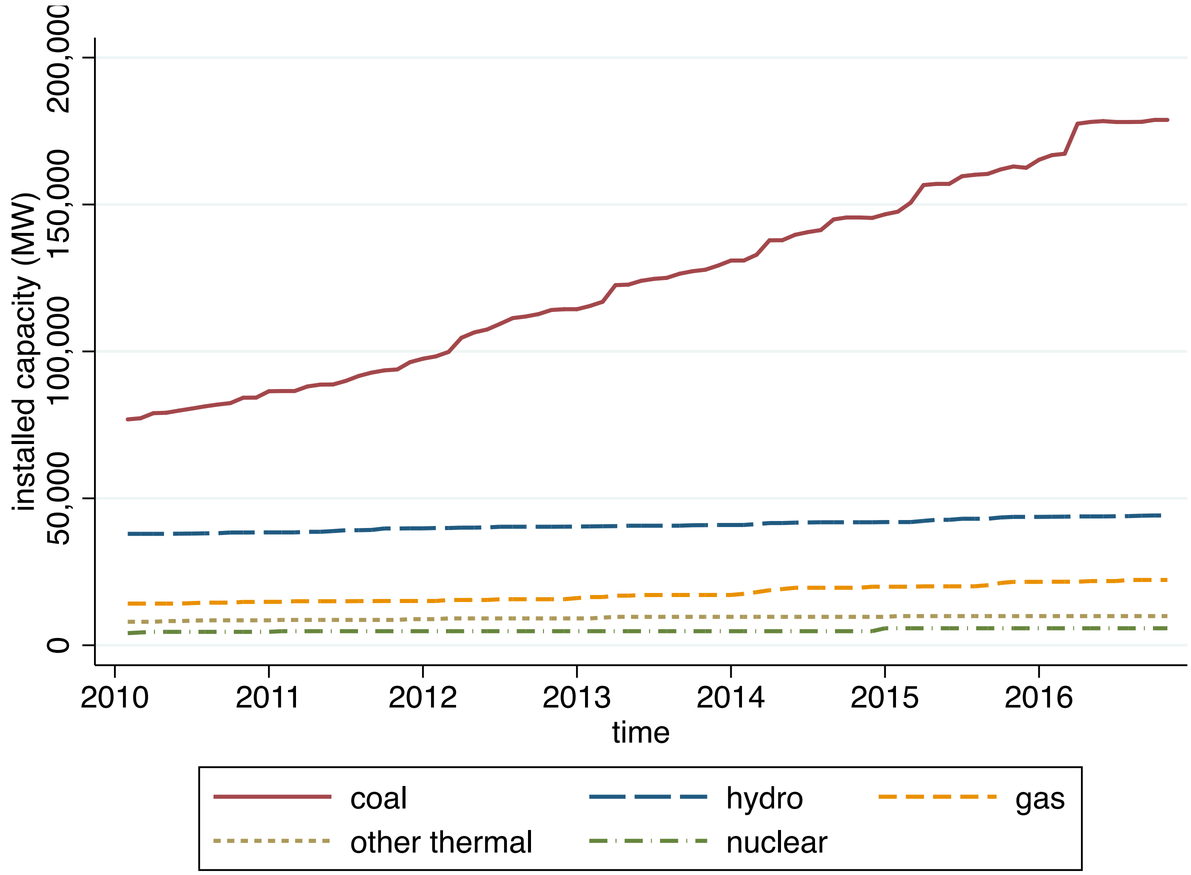


Figure shows month-by-year variation in installed capacity from different fuel sources.

Figure 1: Installed electricity-generating capacity in India 2010-2016

villages that never end up near coal plants. I also explore heterogeneity by socio-economic status and provide evidence that children from both rich and poor households show similar height deficits from exposure to coal plants. This is in contrast to other contexts, where effects are larger for poorer households (Jayachandran, 2009).

Because of data constraints, a growing literature documents the child health impacts associated with polluting activities, such as agricultural fires (Rangel and Vogl, 2016) and forest fires (Jayachandran, 2009), which are easily observable and contribute to air pollution. Coal plants represent another source of variation in air pollution that the literature has begun to explore. In developing countries, Gupta and Spears (2017) study 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. 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.

As an indicator for 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 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 India, and in the context of coal plants, wealth does not appear to be protective against air pollution. This finding highlights an important role for public policy, since even economic development does not mitigate the detrimental effects associated with coal plants.

The paper proceeds as follows. Section 2 summarizes the datasets used for this analysis. 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. Finally, Section 7 concludes.

2 Data

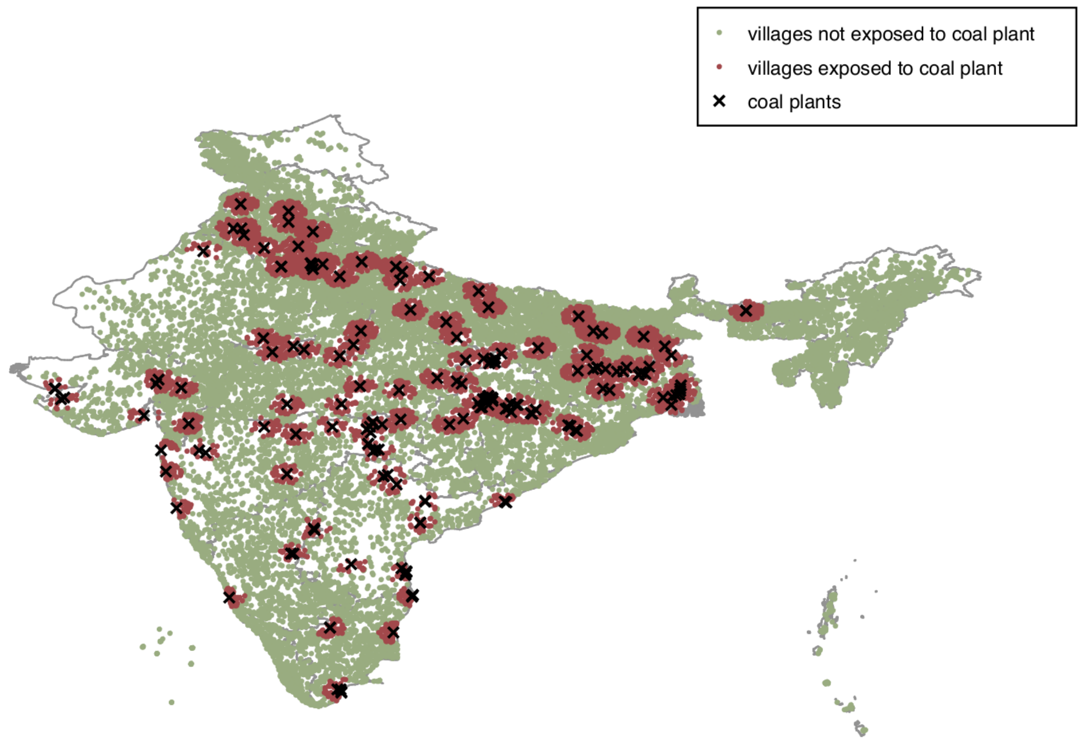
This research uses data on child health from India's Demograhic and Health Survey (DHS) 2015-2016, a dataset on power plant openings and capacity in India from 1922 to 2018, and data on $PM_{2.5}$, particulate matter smaller than 2.5 microns in diameter, estimated from satellite measurements of aerosol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

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. The independent variable of interest is coal plant capacity. 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 kilometers of a coal plant are considered exposed to the plant. Villages that are more than 50 kilometers from all coal plants are considered not exposed.[1] The 50 kilometer 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,[2] I use data estimated from satellite measurements. Specifically, I

[1]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 kilometers from the actual location. Rural clusters are displaced between zero and five kilometers, with 1 percent of rural clusters displaced between zero and ten kilometers. 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.

[2]There are, in fact, no air quality monitors in rural areas.



A village is classified as exposed if it is within 50 kilometers of any coal plant that opens prior to December 2016. A village is classified as unexposed if it is more than 50 kilometers from every coal plant that opens prior to December 2016.

Figure 2: Geography of coal plants and villages

use data collected using the Multiangle Imaging SpecroRadiometer (MISR) V22 aerosol optical depth product, at 17.6 km $\times$ 17.6 km spatial resolution, to estimate $PM_{2.5}$. 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, and are presented in Dey et al. (2012). District-level statistics were extracted using the shape files of the district boundaries in ArcGIS. For further description on how these estimates are constructed, see 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.[3] 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 $PM_{2.5}$ 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 generalized difference-in-difference estimation strategy to identify the effect of coal plant exposure on child height. I include district and time fixed effects to control for variation in capacity and child height over space and secular changes over time. 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.[4] 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 over time 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, and t indexes time of birth 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 kilometers in the month of birth. In regressions using capacity, this

[3]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 $PM_{2.5}$ over time.

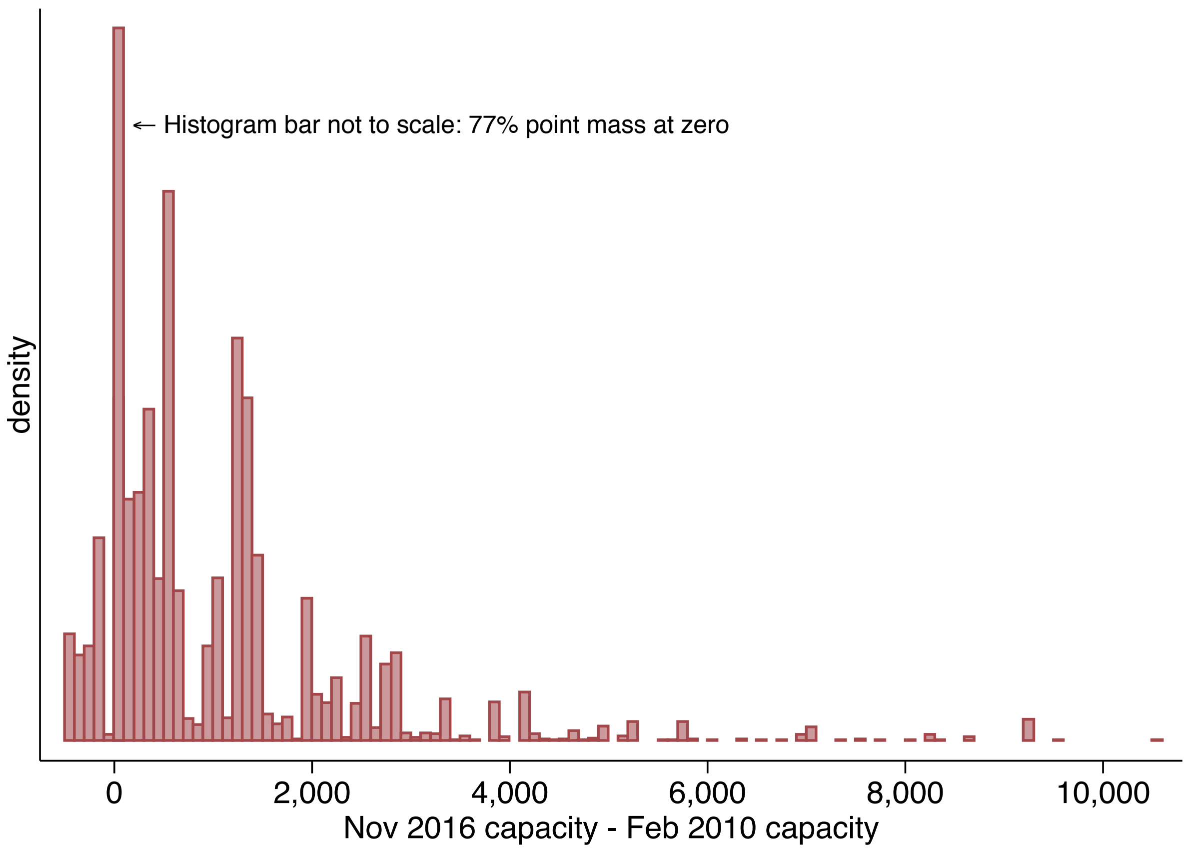
[4]In India, and in other developing countries where environmental risks such as open defecation are particularly severe, height-for-age is decreasing in age because height reflects the accumulating impact of early-life health insults on a child's growth (Victora et al., 2010).

variable has been rescaled so that one unit of capacity represents one median-sized coal plant, which has 1,000 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 time 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.[5] 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, and household open defecation, use of solid fuels for cooking, and electricity access, variables that are indicative of socio-economic status. 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 work to local residents, which made households richer and improved child health in other ways. This violation would, however, bias impacts in the direction of health improvements, and would attenuate any negative impacts. 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 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

[5] 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 obervations of children measured at the same age (in months), born at different periods of time (in month-years).

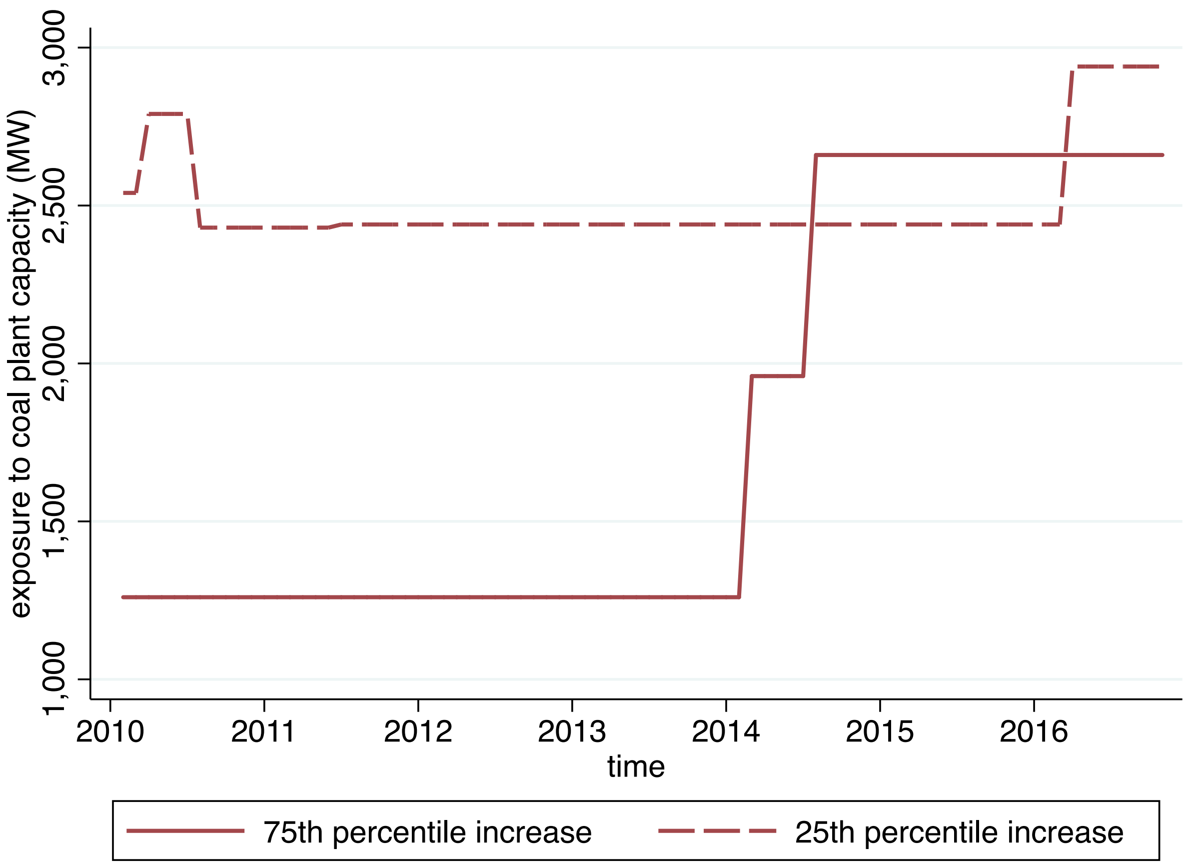


Observations are villages.

Figure 3: Identifying variation: distribution of change in coal plant capacity over time across all DHS villages

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



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.

Figure 4: Increases in coal plant exposure in select villages

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 6, 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). In supplementary analyses presented in Section 4.3, I am able to reject effects on height of exposure after 18 months of age.

Table 1: Summary Statistics

	no exposure (> 50km from coal plant)	**exposure** (≤50km from coal plant)	**difference**
height-for-age	-1.480	-1.489	-0.00900
capacity (GW)	0	1.192	1.192**
units	0	5.493	5.493**
child's age (months)	29.94	30.30	0.357**
female	0.482	0.479	-0.00241
birth order	2.188	2.179	-0.00915
multiple birth	0.0145	0.0130	-0.00151
mom's age at birth (years)	24.32	24.25	-0.0657
institutional delivery	0.802	0.778	-0.0245**
c-section delivery	0.171	0.176	0.00462
mom's height (cm)	151.8	151.4	-0.338**
mom's literacy	0.656	0.671	0.0153**
Hindu	0.795	0.771	-0.0238**
scheduled caste	0.219	0.241	0.0224**
scheduled tribe	0.123	0.0774	-0.0457**
rural	0.762	0.643	-0.120**
defecates in open	0.495	0.419	-0.0763**
uses solid fuel	0.66	0.583	-0.0769**
early breastfeeding	0.695	0.672	-0.0227**
iron supplements in pregnancy	0.781	0.786	0.00564
antihelmintics in pregnancy	0.184	0.174	-0.00977*
n (children under 60 months)	160,282	63,695	

Means are calculated using sampling weights. Standard errors clustered by village. ** p<0.01, * p<0.05, + p<0.10. 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.

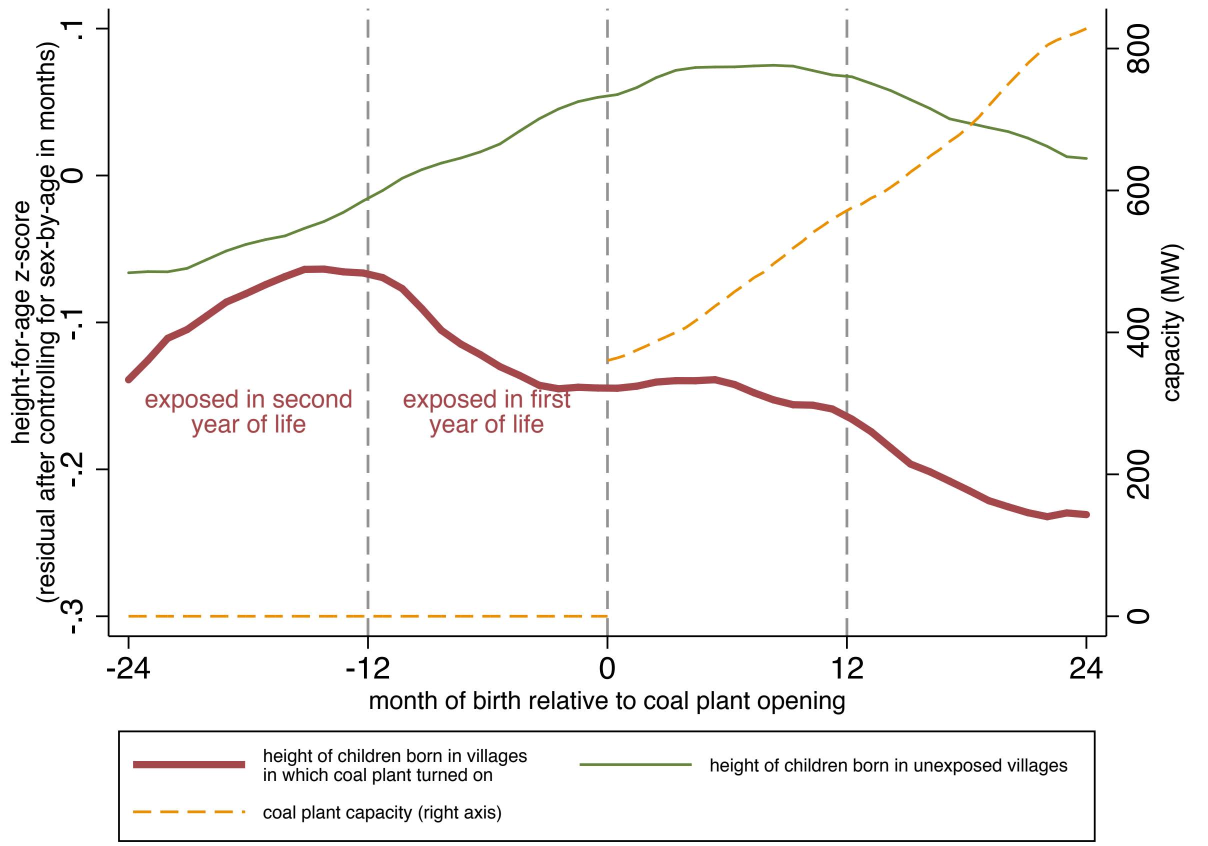
4 Results

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.

The patterns of child height that motivate this analysis are shown in Figure 5, which displays local polynomial regressions of capacity and child height, residualized after controlling for age-by-sex fixed effects. The horizontal axis in this figure is month of birth, relative to a coal plant opening. The thick red line plots the average height of children born in villages in which a coal plant opened nearby. Only children born in villages that started the period under study with zero coal plant capacity and ended the period of study with positive coal plant capacity are included in this line. The dashed yellow line shows the associated coal plant capacity in these villages. On average, the height of children born from 24 months to 12 months before the coal plant opens is increasing over time. However, starting from 12 months prior to the coal plant opening, average height starts decreasing over time. This pattern of child height is in stark contrast to the pattern shown in the thin green line, which represents the average height of children born in unexposed villages. In order to plot the path of child height in unexposed villages on the same x-axis, each unexposed village is randomly assigned a coal plant opening date. In unexposed villages, child height is increasing except for the last 15 months, approximately, while in villages near new coal plants, child height decreases starting 12 months prior to opening.[6] The patterns shown in this figure provide suggestive evidence that exposure to coal plants may be important for child height, and that exposure *in utero* and in the first year of life matter more than exposure after the 15th or 16th month of life.

[6]It is peculiar that average height decreases at all in unexposed villages. This may be an artifact of the way that the DHS was conducted. The survey collected data for a period of two years from January 2015 through December 2016. Children born later in this period would have only been observed in households that were visited later in the survey. Visiting relatively poorer places later in the survey could produce this result. In fact, children born in the last six months of the survey were born to poorer households than children born prior to that period. It is worth noting here that date of birth does not directly map onto the x-axis, which displays time relative to a coal plant opening.



Only children born in villages that started the period under study with zero coal plant capacity and ended the period of study with positive coal plant capacity are included in the thick red line, documenting the height trajectory for children born in villages in which a coal plant turned on. Only children born in villages that were never exposed to a coal plant are included in the thin green line. There is no overlap in children in these two categories. Unexposed villages are assigned a random date of coal plant opening in order to plot the height trajectory of children in these villages on the same x-axis.

Figure 5: Children exposed to coal plants in early life are shorter than unexposed children, and children exposed in later life

4.2 Main result: coal plants predict child height

Table 2 shows the main results of the analysis. In each model, the dependent variable is height-for-age z-score. In panel A, the independent variable is coal plant capacity. A one unit increase in capacity corresponds to an increase in 1,000 MW, or 1 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

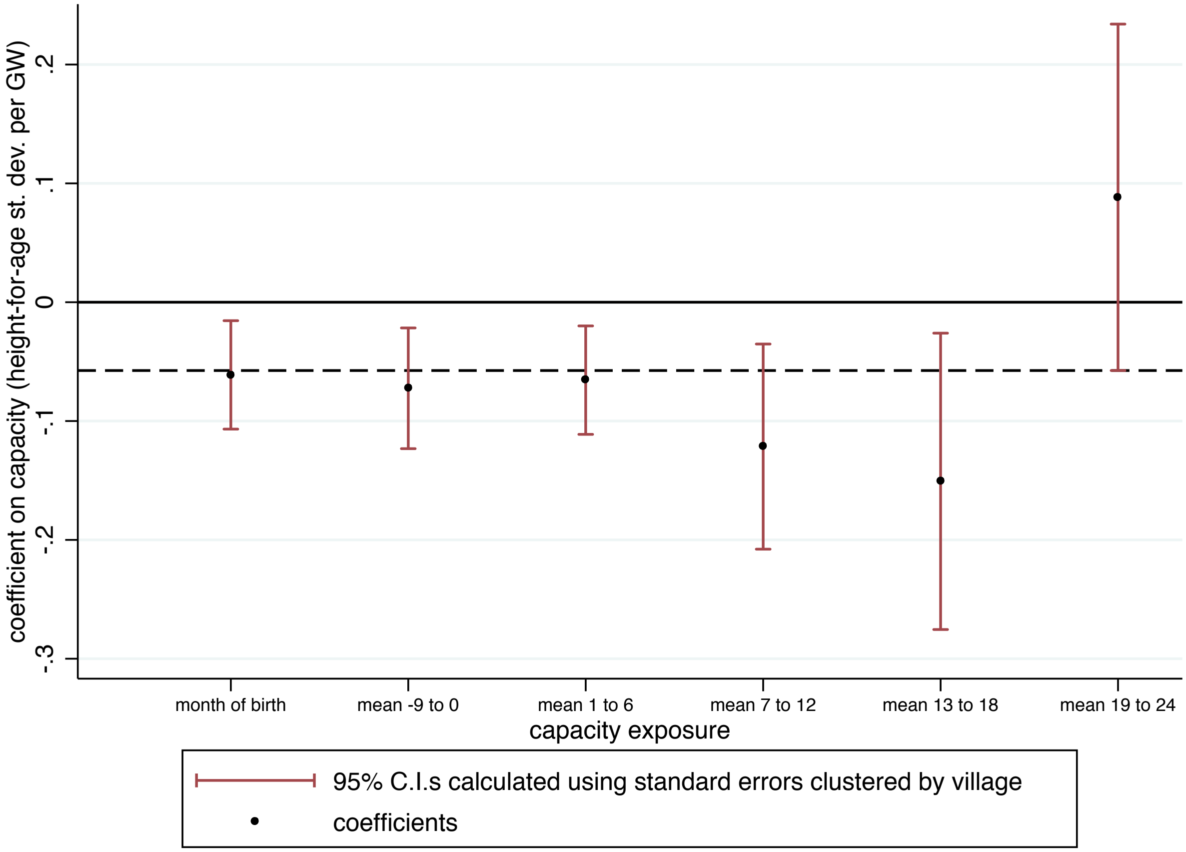
Table 2: Main result

dependent variable:	height-for-age z-score				
	(1)	(2)	(3)	(4)	(5)
Panel A: Coal capacity					
capacity (GW)	-0.104**	-0.103**	-0.0986**	-0.0938**	-0.106**
	(0.0187)	(0.0187)	(0.0190)	(0.0191)	(0.0225)
Panel B: Coal units					
units	-0.0292**	-0.0285**	-0.0276**	-0.0259**	-0.0295**
	(0.00811)	(0.00808)	(0.00808)	(0.00812)	(0.0101)
n (children under 60 months)	223,166	222,619	213,605	213,605	149,680
sex-by-age in months FE	X	X	X	X	X
month-by-year FE	X	X	X	X	X
village FE	X	X	X	X	X
birth characteristics		X	X	X	X
household characteristics			X	X	X
time-by-lat and time-by-long trends				X	
birth characteristics (last birth)					X

Standard errors clustered by village. ** $p<0.01$, * $p<0.05$, + $p<0.10$.

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, 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, and household open defecation, solid fuel use, and electricity. In column 4, I allow for geography-specific time trends by including latitude-by-time and longitude-by-time trends.[7] Geography-specific time trends allow different parts of India to have different time trends. Finally, in column 5, I exclude the time trends, and include iron supplementation and anti-helmintics during pregnancy. Since these variables were only available for the mother's

[7]Latitude, longitude, and time in month-years are continuous variables in these interactions. An alternative test for differential time trends allows states to have differential time trends. In a model that builds off of the specification in column 3 by adding state-by-time fixed effects, the coefficient on *capacity* is -0.0511 (standard error = 0.0210, $t-statistic$ = -2.43). This coefficient is roughly half the size of the coefficients in other models, but it is still 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.



The figure shows coefficients from separate regressions of height-for-age z-score on mean capacity during the described time periods, relative to birth. The dashed line indicates the lowest point of the 95 percent confidence interval for the coefficient on exposure in the 19 to 24 month range. Regressions include age-by-sex, month-by-year of birth, and village fixed effects. 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.

Figure 6: Exposure to coal plant capacity after 18 months does not predict child height deficits

last birth, these regressions have smaller sample sizes than the other models.

4.3 Age of exposure

Figure 5 suggests that exposure to coal plants *in utero* and in early life are important for child height. This section explores the timing of exposure, to the extent the data allow. Figure 6 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:

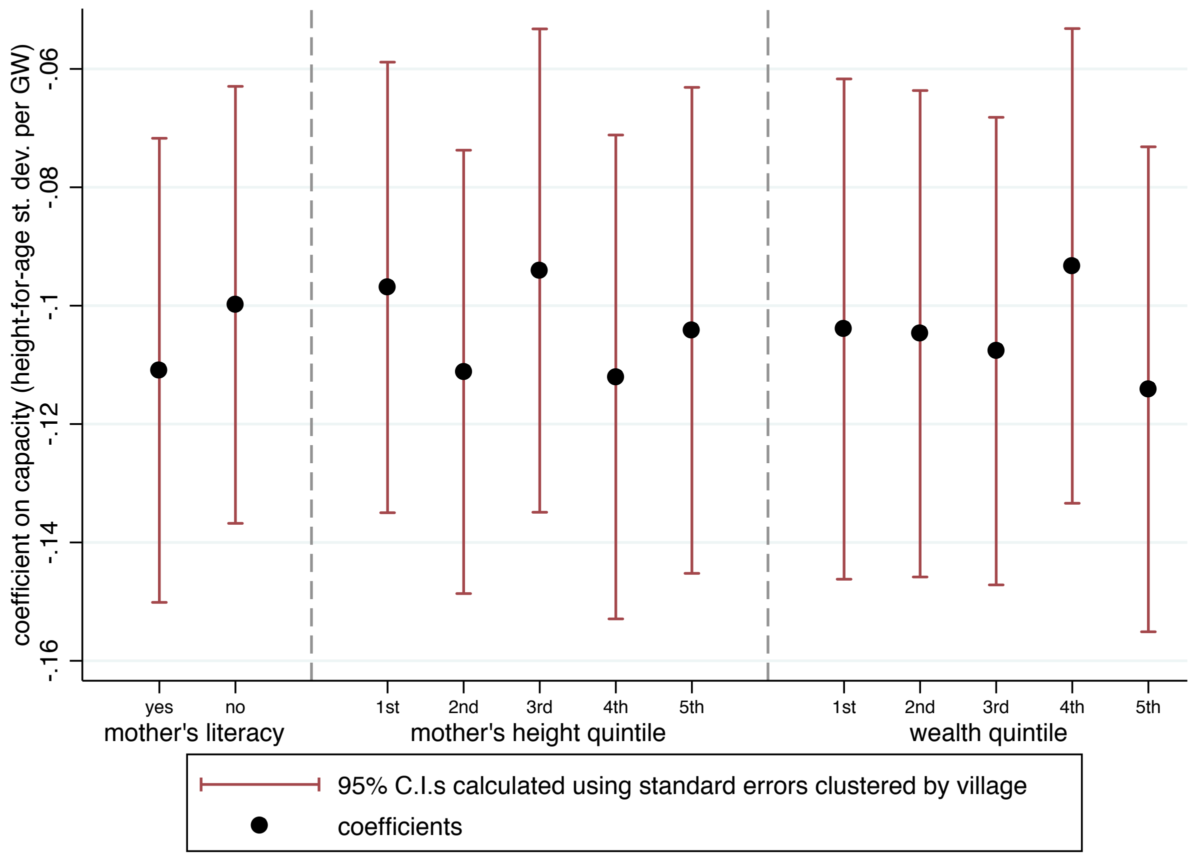
$$height_{ivt} = \beta coalexposure_{vt} + \boldsymbol{B_{ivt}\delta} + \boldsymbol{\alpha_t} + \boldsymbol{\mu_v} + \epsilon_{ivt} \tag{2}$$

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.[8] The vector of birth characteristics, B, represents indicators for age-by-sex categories.

Figure 6 plots coefficients and confidence intervals. It shows that from the *in utero* period to 18 months of age, the effect of exposure is not statistically differentiable from the month of birth. The confidence intervals on mean capacity during the various age ranges until 18 months include the coefficient of the effect of capacity on child height in the month of birth. In contrast, the coefficient on mean exposure in months 19 to 24 is positive and not statistically different from zero. Moreover, the 95 percent confidence interval on this coefficient does not include the coefficients on exposure in earlier time periods. 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, does not affect child height and does not contribute to additional growth faltering.

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 6) 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 intertemporal 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.

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



Differential effects are estimated for mother's literacy, mother's height, and wealth, in three separate regressions, by interacting capacity with indicator variables for the child meeting the specified criteria. Regressions include age-by-sex, month-by-year of birth, and village fixed effects.

Figure 7: No heterogeneous effects by socio-economic status

4.4 Heterogeneity by socio-economic status

Is the effect of coal plant capacity on child height different for children from high socio-economic status households compared to low status households? Effects could be larger for children in poorer families because pregnant moms from these households may be more likely to spend time outdoors working, 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 on cohort size in areas with lower food consumption compared to areas with higher food consumption. These differential effects remain even after controlling for indicators of deprivation, and are an open question for further research.

Figure 7 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.[9] These variables were chosen because they provide an indication of household socioeconomic status.[10] 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:

$$\begin{aligned} height_{ivt} = & \sum_{n=1}^{N} \beta_n capacity_{vt} \times \mathbb{1}[M_{ivt} = value_n] + \\ & \sum_{n=1}^{N-1} \gamma_n \mathbb{1}[M_{ivt} = value_n] + \\ & \boldsymbol{B_{ivt}\delta} + \boldsymbol{\alpha_t} + \boldsymbol{\mu_v} + \epsilon_{ivt} \end{aligned} \tag{3}$$

where $height$, i, v, t, α and μ are as described in Equation 1. n indexes values that 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 7 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. These findings suggest that higher socioeconomic status households are not able to protect their children from air pollution. This differs from the finding in Indonesia, but is consistent with research in India highlighting the difficulty in creating clean air spaces even in very wealthy urban neighborhoods, because of poor infrastructure quality (Vyas, Srivastav and Spears, 2016).

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

[9]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.

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

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 $PM_{2.5}$.[11] 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 8. This figure plots the regression coefficients from a single regression of height-for-age on capacity within different distance bins, from less than 20 kilometers, to between 60 to 70 kilometers.[12] Coefficients are produced from the following regression:

$$\begin{aligned} height_{ivt} =& \beta_1 capacitywithin20km_{vt} + \beta_2 capacity20-30km_{vt} + \\ & \beta_3 capacity30-40km_{vt} + \beta_4 capacity40-50km_{vt} + \\ & \beta_5 capacity50-60km_{vt} + \beta_6 capacity60-70km_{vt} + \\ & \boldsymbol{B_{ivt}\delta} + \boldsymbol{\alpha_t} + \boldsymbol{\mu_v} + \epsilon_{ivt} \end{aligned} \tag{4}$$

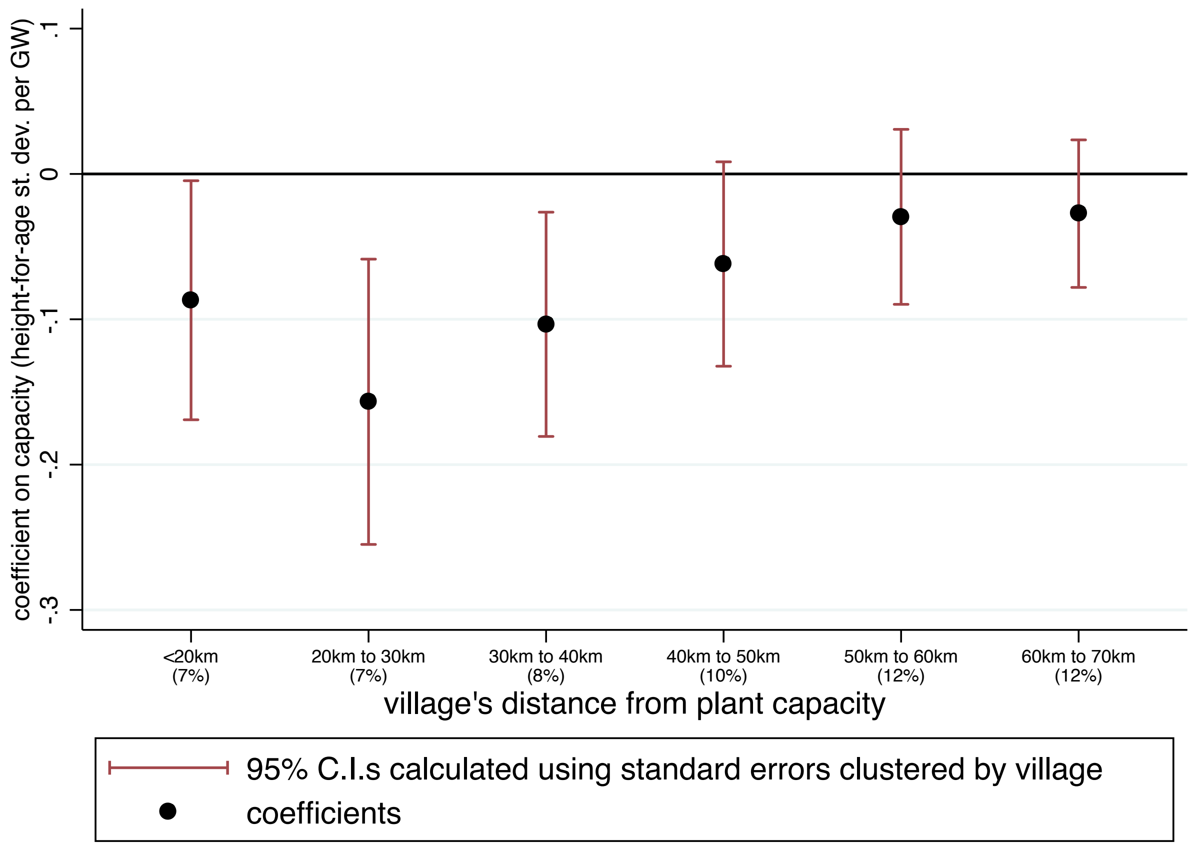
where $height$, i, v, t, α and μ are as described in Equation 1. The vector of birth characteristics, B, represents indicators for age-by-sex categories.

Except for the first distance bin, the effect size of coal plant capacity becomes closer to zero as distance from the coal plant increases.[13] The effects of capacity within bins less than 40 kilometers away are statistically different from zero at the 5 percent level, the effect within the 40 to 50 kilometer bin is statistically different from zero at the 10 percent level, and beyond 50 kilometers, the effect of capacity is not statistically different from zero. This finding lends credence to the 50-kilometer cutoff used in the main results, since exposure beyond this distance does not statistically significantly influence child height. Table A2

[11] 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.

[12] Within 10 kilometers, and between 10 and 20 kilometers, are combined into one single bin because of sample size constraints.

[13] 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 = 1.07, p-value = 0.30).



Coefficients from a single regression of height-for-age z-score on capacity within each of the described distance bins. Regression includes age-by-sex, month-by-year of birth, and village fixed effects. Numbers in parentheses are the fraction of the sample that have positive capacity within the distance bin.

Figure 8: Effect of coal capacity on height attenuates with distance

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.[14]

[14]Given these coefficients, the effect of capacity reaches zero at about 80 kilometers distance from the plant. This is considerably larger than the 50 kilometer 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 to which the village is exposed, where weights are the fraction of total capacity that the coal plant contributes for that village in that month.

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. For instance, if a new coal plant, or an expansion of an existing one, brought jobs which provided higher incomes to local families that allowed parents to create a household environment more conducive to child development, then the true effect of coal plants on child height would be larger than the effects that I find.

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

dependent variable:	**mom's age at birth**	**birth order**	**multiple birth**	**institutional delivery**
	(1)	(2)	(3)	(4)
capacity (GW)	0.0223	0.00710	-0.000792	0.00306
	(0.0412)	(0.0119)	(0.00234)	(0.00362)
n	223,166	223,166	223,166	222,620
dependent variable:	**c-section**	**breastfeeding w/in 1 hr of birth**	**iron supplements**	**drug for intestinal parasites**
	(5)	(6)	(7)	(8)
capacity (GW)	0.000322	-0.000924	-0.00298	0.00157
	(0.00322)	(0.00511)	(0.00411)	(0.00453)
n	223,166	158,645	165,592	164,451

Regressions include age-by-sex, month-by-year of birth, and village fixed effects. Standard errors clustered by village. ** p<0.01, * p<0.05, + p<0.10.

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 by 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 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 4.3 cannot rule out that exposure until 18 months of age is relevant for child height.

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
	(0.00264)
continuous time (cmc)	0.0365+
	(0.0208)
n	159,716
Panel B:	
reference: effect of coal capacity	-0.104**
(table 2, column 1)	(0.0187)

Sample in Panel A consists of children in villages that had no exposure to coal plants by the end of DHS data collection (Dec 2016). $futureplant$ is an indicator for getting exposed to a plant by March 2018. Regression includes age-by-sex, month-by-year of birth, and village fixed effects. Standard errors clustered by village. ** p<0.01, * p<0.05, + p<0.10.

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:

$$height_{ivt} = \beta futureplant_v \times time_{ivt} + \eta time_{ivt} + \boldsymbol{B_{ivt}\delta} + \boldsymbol{\mu_v} + \epsilon_{ivt} \tag{5}$$

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 time 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 over time. Panel B shows the main effect estimated from Table 2, column 1, for comparison. The differential time 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 Other robustness checks

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 A3. 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.[15] I also implement a Box-Cox

[15] 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 5 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

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 A4 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.

Finally, Table A5 presents results from regressions which cluster errors in different ways (Cameron and Miller, 2015): at the village, district, and plant levels. It also explores two-way clustering. As expected, standard errors become larger when errors are permitted to be correlated across larger geographic areas. Notably, however, every type of error correlation tested finds significant effects of coal plant capacity on child height.

7 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 around 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 corre-

a transformation that is defined at zero, the inverse hyperbolic sine function. In both of these regressions, coefficients on the variables of interest are statistically significant, 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.

lated, 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 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.

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.

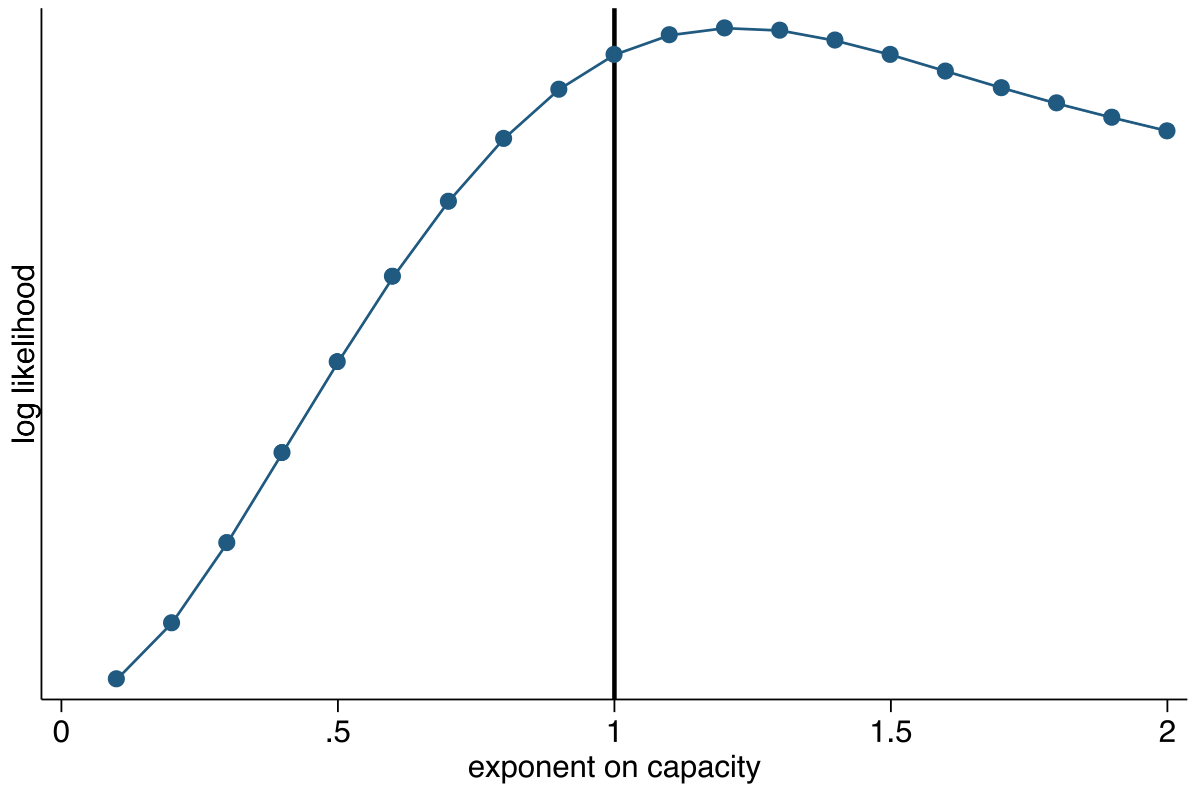
Spears, Dean. 2018. "Exposure to open defecation can account for the Indian enigma of child height." *Journal of Development Economics*.

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.

Appendix



Note: Figure displays log likelihoods for separate regressions, each with a different exponent on capacity. Each regression includes age-by-sex, time, and village fixed effects.

Figure A1: Effects are potentially steeper at higher capacity levels

Table A1: Coal plant capacity is associated with higher ambient air pollution

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-months)	43,820	43,808	43,808
district FE	N	Y	Y
month-by-year FE	N	N	Y

Standard errors clustered by district. ** p<0.01, * p<0.05, + p<0.10. Sample sizes vary because some fixed effects categories lack within-category variation in the independent variable (resulting in that category being dropped).

Table A2: Effect of coal capacity on height by distance

dependent variable:	**height-for-age z-score**
capacity (GW) X distance (km)	0.00169*
	(0.000842)
capacity (GW)	-0.137**
	(0.0410)
distance (km)	0.00115
	(0.000731)
n	94,481

Sample consists of children born in villages ≤70 km away from a coal plant at some point in time during the study period. *capacity* is the total capacity that the village in which the child is born is exposed to in the month of birth. 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 that month. Regression includes age-by-sex, month-by-year of birth, and village fixed effects. Standard errors clustered by village. ** p<0.01, * p<0.05, + p<0.10.

Table A3: 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*
	(0.0187)		(0.0263)			(0.0476)
capacity x 1[1st quartile]		0.123				
		(0.125)				
capacity x 1[2nd quartile]		-0.108+				
		(0.0563)				
capacity x 1[3rd quartile]		-0.0587				
		(0.0359)				
capacity x 1[4th quartile]		-0.0964**				
		(0.0195)				
capacity^2			-0.00322			
			(0.00241)			
ln(capacity)				-0.0249*		
				(0.00988)		
sinh^{-1}(capacity)					-0.163**	
					(0.0359)	
above median spline						-0.00386
						(0.0565)
n	223,166	224,188	223,166	224,188	223,166	223,166
F-statistic $\beta^{\text{1st q}} = \beta^{\text{2nd q}} = \beta^{\text{3rd q}} = \beta^{\text{4th q}}$		2.483				
P-value		0.0589				

Regressions include age-by-sex, month-by-year, and village fixed effects. *capacity* = 0.01 replaces *capacity* = 0 for regression using $ln(capacity)$. Standard errors clustered by village. ** p<0.01, * p<0.05, + p<0.10.

Table A4: The main effect is not driven by outliers

	dependent variable: height-for-age z-score					
sample:	**born in place of interview**	**exposed villages only**	**children exposed to capacity ≤**			
			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*
	(0.0193)	(0.0203)	(0.0209)	(0.0246)	(0.0315)	(0.0397)
n	206,053	63,450	222,157	220,234	218,383	209,578

Regressions include age-by-sex, month-by-year of birth, and village fixed effects. Standard errors clustered by village. ** p<0.01, * p<0.05, + p<0.10.

Table A5: Level of clustering

	type	first dimension		second	number	standard	
		exposed	**unexposed**	**dimension**	**of clusters**	**error**	**t-stat**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A	one-way	village		none	28,164	(0.0187)	-5.525
B	one-way	district		none	640	(0.0294)	-3.522
C	one-way	nearest plant	village	none	28,341	(0.0214)	-4.843
D	one-way	nearest plant	district	none	817	(0.0220)	-4.713
E	one-way	nearest plant	state	none	213	(0.0244)	-4.239
F	one-way	nearest plant	single	none	178	(0.0214)	-4.831
G	two-way	nearest plant	single	district		(0.0298)	-3.478
H	two-way	nearest plant	single	village		(0.0214)	-4.831

all models:

dependent variable:	height-for-age z-score
point estimate:	-0.104
n:	223,166

All regressions include age-by-sex, month-by-year of birth, and village fixed effects. Each row displays results from a separate regression, which clusters errors in a different way.