**Research Article**

# Direct evidence of the impact of early-life exposure to ambient $PM_{2.5}$ air pollution on later-childhood height-for-age in India

Dean Spears[1,2], Sagnik Dey[3], Sourangsu Chowdhury[3], Sangita Vyas[1], and Joshua Apte[4]

1. Department of Economics and Population Research Center, University of Texas at Austin, USA

2. Economics and Planning Unit, Indian Statistical Institute – Delhi Centre, India

3. Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, India

4. Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, USA

**The disease burden from ambient fine particulate ($PM_{2.5}$) exposure in India has been estimated so far using risk functions based on studies done elsewhere. Here we provide the first direct evidence of the impact of ambient $PM_{2.5}$ exposure on child health in India using measurements from nationally-representative anthropometric data matched to satellite-based exposure data. We apply fixed effect regression with child height-for-age as the dependent variable and district-month-level early-life exposure to ambient $PM_{2.5}$ as the independent variable. We show that a 100 $\mu g/m^3$ increase in ambient $PM_{2.5}$ exposure leads to a 0.05 standard deviation decrease in height-for-age after controlling for district-specific seasonality, household properties, and other confounding factors. We find effects on both rural and urban children, and cannot reject that the shape of the concentration-response curve is linear. Because average exposure to ambient particulate pollution is high in India, our results recommend ambient air pollution as public health policy priority.**

Ambient and household $PM_{2.5}$ exposure have been linked causally to various child health outcomes such as lower respiratory infection (LRI)[1], sudden infant death syndrome[2], low birth weight [3,4], intrauterine growth retardation[5,6] and reduced size[7]. The recent Disease Burden of India (DBI) study[8] attributes 5.1 (4.1-6.3) million disability adjusted life years (DALY) and 0.06 (0.04-0.07) million deaths of children (<5 years) to LRI, due to ambient $PM_{2.5}$ exposure in India. The child mortality burden due to household $PM_{2.5}$ exposure in India is estimated to be equally large at 0.05 (0.03-0.06) million. However, these estimates relied on exposure-risk functions that were developed from epidemiological studies carried out in developed countries. Direct evidence of the impact of ambient $PM_{2.5}$ exposure on child health in India is lacking so far.

One widely-studied marker of early-life health insults in India is the average height of children[9]. Children in India are unusually short in international comparison, on average. Many correlates and causes of India's mean child height deficit (or stunting), a clinically extreme height defecit, have been documented in the demographic, epidemiological, and econometric literatures[10-12]. Intra-household exposure to particulate matter has also been found to impact child growth. Prevalence of stunting was significantly higher (relative risk ratio RRR = 1.84, 95% uncertainty interval (UI) 1.44-2.36) amongst children living in the households that use solid fuel compared to the children living in the households using clean fuel[5]. Using data from the 2005-2006 National Family Health Survey (NFHS), a recent study[6] has shown strong evidence that household solid fuel exposure increases the risk of stunting and reduces the height-for-age measure of the Indian children. More recently, ambient air pollution is also identified as a factor to impact child growth. Another study[7] observed significant increases in the relative risk of child stunting and wasting in Bangladesh associated with higher levels of in utero exposure to ambient air pollution. Ambient $PM_{2.5}$ exposure in India is quite large and varies in the range 3.7 to 148 $\mu g/m^3$ at annual scale [8,9]. To our knowledge, no study has ever been carried out to examine the impact of early-age ambient $PM_{2.5}$ exposure on child height-for-age in India.

To address this question in a sample representative of the population of children under five in India – a population exposed to a large range of ambient $PM_{2.5}$ – here we report an observational analysis of India's 2015-2016 Demographic and Health Survey (DHS), matched to district-month level air pollution, as measured by satellite remote sensing. The association between child height-for-age and early-life exposure to air pollution is estimated using a fixed effects econometric strategy that accounts for fixed differences across villages, for secular trends over time, and for district-specific seasonal patterns. The resulting association between childheight outcomes and exposure to ambient $PM_{2.5}$ is identified off of unpredictable deviations from these trends. We show that ambient $PM_{2.5}$ exposure reduces child growth in India across rural and urban areas. The rate of decrease of child height-for-age with an increase in ambient $PM_{2.5}$ exposure at their month of births is consistently similar across the seasons and other confounding factors. We also show evidence for a linear shape of the exposure-response function. These results are the first direct confirmation of the impact of early-age ambient $PM_{2.5}$ exposure on child growth in India (or anywhere in the world). We hope that our results might help formulate policy to curb ambient $PM_{2.5}$ exposure at regional scale.

## Results

**Summary and descriptive statistics.** Figure 1 describes the study population and sample. Height is measured for 225,002 children under 5 in the DHS. We are able to match air pollution data to children born before January 2016, who are 97% of those with measured height, resulting in a final sample of 218,152 children. Summary statistics for these children are presented in Table 1. As Table 1 shows, the DHS is also a rich source of further information on children and their households, which we exploit in robustness checks as regression controls.Table 1 presents sample means that summarize our data. By separating the sample by quintiles of ambient $PM_{2.5}$ exposure, the table describes the correlates of the independent variable, and therefore some potential omitted variable bias threats. Children who are exposed to higher ambient $PM_{2.5}$ also tend to be disadvantaged in other ways. They come from larger families, have shorter mothers, live in households that are more likely to defecate in the open, use traditional fuel, and live in the poorer northern plains states of Uttar Pradesh and Bihar. However, much of these correlations are absorbed by our controls and by our primary sampling units (PSU; Figure 1) and seasonality fixed effects.

**Non-parametric analysis.** Before proceeding to our main regression results, we use non-parametric methods to illustrate the relationship between ambient $PM_{2.5}$ exposure and subsequent child height. Figures 2, 3, and 4 use locally-weighted polynomial regression; all three show evidence of a robust association. Figure 2 reveals a negative gradient between ambient $PM_{2.5}$ exposure and child height for both rural and urban children. Overall, rural children are shorter, on average, because they are more exposed to other factors associated with growth faltering[10,11]; this is visible in the fact that the rural line is below the urban line. The principal results of Figure 2, however, are that both lines have an apparently linear downwards gradient, and that they are parallel, which is consistent with a comparable gradient for rural and urban children.

The seasonality of exposure to ambient $PM_{2.5}$ in India is reflected in Figure 3. Here, each season-of-birth is plotted separately, to ensure that season of birth is not a biasing omitted variable. Ambient $PM_{2.5}$ levels reach the highest levels in the winter months of November through January, which is visible in the fact that this line extends the furthest to the right. However, Figure 3 provides evidence that the pattern in Figure 2 does not merely reflect a seasonal trend in height, because similarly-steep downward gradients are present in all four seasons. In other words, the season of birth may matter for a child's outcomes, but it is not a confounder in the gradient that we document.

Finally, by splitting the sample by decile of mother's height, Figure 4 speaks to the possibility that the gradient on Figure 2 merely reflects confounding heterogeneity across household environments or among children's genetic endowments. Each line is separately computed for children of mothers with similar heights. The lines are approximately parallel: within each decile, children exposed to more ambient $PM_{2.5}$ in their month of birth are shorter, on average. Mothers' height (and its correlates) does not appear to be a potentially biasing omitted variable for our results.

**Effect of month-of-birth exposure.** Table 2 presents our main results: fixed effect regression results following equation 1 (see Methods for more details). Ambient $PM_{2.5}$ exposure is divided by 100 for ease of interpretation of the coefficients. Across the alternative specifications in columns 1 through 5, a 100 $\mu$g/m$^3$ increase in ambient $PM_{2.5}$ exposure is associated with an approximately 0.05 standard deviation decrease in child height-for-age. Column 4 verifies that the result is unchanged after controlling for household fuel use. Column 5 is a falsification test. Ambient $PM_{2.5}$ exposure two years before the child is born does not predict height and does not change the coefficient of interest. Columns 7 and 8 find similar results when state-month or PSU-month fixed effects for seasonality are used instead of district-months.

Although not reported in the table, we conducted a further robustness check that our result is not driven by the January height pattern documented in the DHS literature. Omitting children born in January results in an essentially unchanged estimate of -0.052 (standard error = 0.023; p = 0.024). Additionally replacing year of birth fixed effects with a larger set of state-specific year of birth fixed effects reduces precision by consuming degrees of freedom, but does not qualitatively change the estimate (-0.069, standard error = 0.030, $p$ = 0.022 in the most fully controlled specification).

Figure 5 presents seven estimates of equation 2 (see the Methods section), each for ambient $PM_{2.5}$ exposure in a separate three-month age range. Only the early-life period at and immediately after birth shows a coefficient that is statistically distinguishable from zero. Most of the other coefficients are close to zero, and none is as large in absolute value as the one for ages 0-2 months. These results are consistent with evidence in the literature[12] that early-life is a critical period for the determination of child height. Some prior literature has documented evidence for effects on health of *in utero* exposure[13]; although we do not detect *in utero* effects, the confidence intervals on pre-birth exposure cannot rule out effects about half as large as the effect that we find for exposure in the first months of life.

**Shape of concentration-response function.** Three tests for non-linear concentration-response functions each fail to reject that a linear shape fits the data. Moreover, each approach suggests that, if anything, effects are steeper at higher concentration levels. Table 2 includes non-linear candidate function shapes. As column 6 shows, a natural log functional form – which would be consistent with the hypothesis that the concentration-response function shows diminishing marginal costs – has a coefficient that is statistically distinguishable from zero, but fits the data less well (as measured by a $t$-statistic) than a linear form. Column 9 includes a linear spline that allows a different slope at above-median levels of ambient $PM_{2.5}$. Although the two $PM_{2.5}$ terms are jointly statistically significant ($F$ = 2.99; $p$ = 0.051), neither is individually statistically significantly different from zero. The coefficient on the spline term is negative, indicating that, although this model does not fit the data better than a simple linear concentration-response function, the sign suggests a steeper concentration-response function at higher levels of exposure.

Finally, Figures 6 and 7 present two further ways of investigating whether the concentration-response function has evidence of a non-linear shape. Figure 6 investigates higher-order polynomials, beyond linear. It presents results for quadratic, cubic, quartic, and quintic polynomial forms. Of these, only the quadratic form is jointly statistically significant, and its fit does not improve on a linear functional form. The coefficient estimates for higher-order polynomials suggest, in each case, effects that are, if anything, steeper at higher levels of exposure. Figure 7 graphically presents the results of a Cox-Box transformation, a standard parameterization of a curved relationship, detailed in the Methods section. It presents log-likelihoods for a range of power transformations of $PM_{2.5}$, estimated with and without the full set of controls. The likelihood of the model is maximized at or just above an exponent of 1, indicating that a linear model (or perhaps one with slightly increasing marginal effects) best fits these data.

## Discussion

This paper reports an ecological analysis of variation in remotely-sensed ambient $PM_{2.5}$ exposure data at the district-month level. Ecological analysis is often used to generate hypotheses for further investigation using more rigorous methods. In this instance, there are inherent limits to the possible study design: ambient air pollution is an important topic of study, but it must vary at a geographic level, and is not amenable to experimental manipulation. Although place, time, and season fixed effects limit the role of residual confounding, we are unable to use an econometric design that exploits a specific, known source of variation in $PM_{2.5}$ exposure, such as a policy change. Nevertheless, measurement error could, in principle, be improved by a study that records child-level exposure to ambient $PM_{2.5}$ with a system of mobile child-level personal monitors.

This is the first direct evidence of the ambient $PM_{2.5}$ exposure impact on child health at a country level. Although child height has traditionally been interpreted as a measure of “malnutrition,” it is increasingly recognized that anthropometric outcomes such as height reflect the totality of early-life net available nutrition, including losses due to diseases [14], and including growth effects of lung function. Our data do not allow us to observe disease directly; indeed our health data reflect only conditions at the time of the survey, and not during the critical period of the child’s birth. However, mechanisms in the literature are consistent with the effect that we document[7]. For example, lung function growth has been linked to children’s exposure to particulate matter[15]. More generally, child growth is highly correlated, at the population level, with early-life mortality, which has been interpreted as a consequence of the role of infectious disease. Mortality is correlated with average child height because survivors’ growth is “scarred” by its early-life disease[16]. This could plausibly include respiratory disease.

Many studies in the air pollution literature use mortality as a dependent variable, typically from census or vital registration data. But India does not have a vital registration system, like many other developing countries. As Setel and colleagues[17] explains: “Most people in Africa and Asia are born and die without leaving a trace in any legal record or official statistic." Therefore, we study height-for-age as a dependent variable, because it is a summary of early-life health that is a continuous variable, and therefore offers high statistical power even in a survey sample, relative to dichotomized outcomes such as stunting or infant mortality [18].

We document an effect of early-life exposure to ambient $PM_{2.5}$ on subsequent height-for-age in later childhood, using India's most recent DHS, which measures the children under five years old of a nationally representative sample of reproductive age women. The effect size we estimate is plausibly small for any one child, but many children are exposed to it. Moreover, ambient $PM_{2.5}$ concentrations in India are higher than World Health Organization guideline. With an effect size of 0.05 associated with a linear difference of 100, the average child in India is about 0.027 height-for-age standard deviations shorter than he or she would be if exposed to very low levels of air pollution, an effect multiplied by almost 30 million births per year. Because the ambient $PM_{2.5}$ exposure is projected to increase in India in near future under climate change scenarios [19], the health burden that we quantify here could potentially increase unless appropriate policy is taken to reduce air pollution throughout India. In particular, although policy conversations often focus on Delhi (and, to a lesser extent, other big cities), we find results throughout India for rural and urban children, suggesting that the policy challenges are significantly broader than is commonly understood. Because child height has lasting consequences for human capital [12,20], this is a problem with ramifications throughout the Indian society and economy.

## Methods

**India's 2015-2016 Demographic and Health Survey.** The dependent variable and regression controls are taken from India's most recent Demographic and Health Survey (hereafter DHS; in India, the DHS is also known as the National Family Health Survey). These data were collected from a nationally-representative sample of women of reproductive age. In particular, the sample was constructed to permit district-level estimates for all 640 districts in India at the time of the 2011 census. These data were collected between January 2015 and November 2016.

Our dependent variable of interest is a child's height-for-age *z*-score, scaled according to the World Health Organization 2006 reference population mean and standard deviation by sex and age-in-months (WHO 2006). In the DHS, height is measured for children less than five years old at the time of the survey. The sex and month of birth (e.g. August 2011) is also recorded for each child with measured height.

**Air pollution data by district-month.** Each child is matched to average ambient $PM_{2.5}$ exposure in his or her district of residence, during the month in which he or she was born. This matching implicitly assumes that the district where children live at the time of the survey is the same as the district where children lived when they were born.

The absence of systematic ground-based $PM_{2.5}$ measurements at desirable spatial resolution prompted us to use satellite-derived $PM_{2.5}$ for this study. We use the Multiangle Imaging SpecroRadiometer (MISR) retrieved daily aerosol optical depth (AOD) V22 product at 17.6 km spatial resolution to estimate the $PM_{2.5}$ with the help of a spatially and temporally varying conversion factor ($\eta$). $\eta$ is derived from of GEOS-Chem chemical transport model simulations and depends on aerosol vertical distribution, emission and meteorological factors like temperature, relative humidity and precipitation. Details about the conversion factor η are discussed elsewhere[9,21,22]. MISR AOD product was earlier extensively evaluated for the Indian subcontinent [23]. The satellite-retrieved $PM_{2.5}$ was bias-corrected using coincident ground-based quality controlled measurements following our earlier study[8,9]. The district-level statistics are extracted using the shape files of the district boundaries in ArcGIS. We generate a monthly $PM_{2.5}$ exposure database for 15 years (2001-2015), although because height is only measured in the DHS for children under 5, no child in our sample was born before 2010.

**Fixed effects econometric strategy.** The central empirical strategy of this paper is fixed effects regression, with child height-for-age as the dependent variable, and early-life district-month-level exposure to ambient $PM_{2.5}$ as the independent variable of interest. The DHS is a cross-sectional survey that measured children under five at different ages; because age is predictably correlated with height-for-age[24], each regression therefore controls for 119 age-in-months-by-sex indicators. Each regression also controls for fixed effects for place, season, and year, as detailed in 'Results' section. We first present the econometric strategy for the main result followed by further investigation of exposure at other ages and robustness checks in which we allow the concentration-response function to take non-linear shapes. All analysis in the paper is computed with Stata 12.1.

**Non-parametric descriptive regression.** Exposure to ambient $PM_{2.5}$ is not randomly allocated across children. This fact raises the possibility that any apparent association between air pollution and child outcomes could, in fact, reflect omitted variables such as seasonality of births or geographic heterogeneity across India. Graphs of non-parametric regressions in split samples can be a method to investigate whether the variable over which the sample is split confounds the relationship of interest[25]. In particular, in Figures 2, 3, and 4 we plot locally-weighted kernel regressions, computed in Stata with the default Epanechnikov kernel function. Each figure uses, as the independent variable, ambient $PM_{2.5}$ in the district month of birth and, as the dependent variable, height-for-age residuals after regression on 119 age-in-months-times-sex fixed effects, to account for the fact that children were measured at different ages.

Figure 2 splits the sample by rural and urban. Although much of the discussion of air pollution in India focuses on urban places, this permits us to see if there is a gradient for rural children. Figure 3 splits the sample by season of birth; within each season, only observations in the 5$^{th}$ to 95$^{th}$ percentiles of $PM_{2.5}$ exposure are plotted (because non-parametric statistics of this sort require an adequate sample size). This split permits us to investigate whether any gradient is present within seasons, and not merely a reflection of omitted seasonality. Finally, in recognition of the fact that $PM_{2.5}$ exposure is high in the poorer regions of India, where other factors also constrain child growth, Figure 5 splits the sample into ten partitions, according to deciles of mothers' height; within each decile of mother's height, only observations in the 5$^{th}$ to 95$^{th}$ percentiles of $PM_{2.5}$ exposure are plotted. If a gradient is visible within each decile, then it is evidence that the association between air pollution exposure and child height is not fully driven by the omitted variables that are correlated with mothers' height.

**Main strategy.** Our main analysis estimates fixed effects regressions of the following form:

$$h_{ipdmy} = \beta x_{dmy} + \boldsymbol{\alpha}_{pd} + \boldsymbol{\gamma}_{dm} + \boldsymbol{\delta}_{y} + X_{ipdmy}\boldsymbol{\theta} + \varepsilon_{ipdmy} \qquad (1)$$

where *i* indexes individual children, *p* places (survey PSUs, such as urban blocks or rural villages), *d* districts, *m* calendar months (such as February), and *y* calendar years, such as 2008. The dependent variable, *h*, is child *i*'s height-for-age *z*-score. The independent variable of interest, $x_{dmy}$, is $PM_{2.5}$ in district *d* in month *m* of year *y*, corresponding to child *i*'s birth month. Fixed effects are $\boldsymbol{\alpha}_{pd}$, 27,266 local places (PSUs); $\boldsymbol{\gamma}_{dm}$, 7,679 categories of district-month (such as for Februarys in Sitapur district, or Aprils in Kanpur district); and $\boldsymbol{\delta}_{y}$, 6 calendar years, to capture any secular time trend. Child-level covariates $X_{ipdmy}$ include the age-by-sex fixed effects and a vector of extended controls.

We add fixed effects and controls in stages to verify that the main effect estimate, $\hat{\beta}$, is robust to respecification. In particular, we first estimate the model without PSU fixed effects. PSU fixed effects would account for any fixed geographic heterogeneity such as market [26], local open defecation [10], or the religious composition of the neighborhood[27]. We then add a set of extended regression controls, intended to control for other known determinants of child height: the height of the child's mother, the child's birth order, the number of siblings born to its mother by the time of the survey, household open defecation, the caste and religion of the child's household [11], the and child's mother's relationship to the head of the household. We further add indicators for the household's cooking fuel type, as a proxy for household $PM_{2.5}$ exposure.

Our main specifications control for district-month fixed effects[28]. This strategy allows each district to have any distinct seasonality pattern, and identifies effects off of deviations from each district's seasonal patterns. Fertility is known to be predicted by seasonal patterns [29], but parents would not be able to make fertility decisions (nine months in advance) based on the realized deviation from seasonal trends. Therefore, predictable seasonality does not confound our results. We include robustness checks with coarser (state-month) and finer (PSU-month) controls for seasonality.

Finally, we conduct several additional robustness tests. As a falsification test, we control for ambient $PM_{2.5}$ in the same district-month, two years before the month of birth; if our identification strategy is credible, this control should not predict height nor change our estimate. Few studies[30,31] both document a pattern in DHS data of a drop in child height of children born in January, relative to children born in December; they note that this pattern could bias studies that identify effects on height from child season of birth. This is not what our work does. We control for seasonality and identify off of deviations from it, and we control non-parametrically for age-in-months-times-sex. However, to verify that this pattern is not a threat to our conclusion, we compute a robustness check omitting children reported to be born in January.

Standard errors are clustered by 640 districts, to permit arbitrary correlation of error terms over space and time within districts [32]. DHS data include sampling weights; although we use weights for our summary statistics in Table 1, sampling weights are not appropriate for estimating relationships [33], so we do not use weights in any of our regression results.

**Age of exposure.** Our main specification only investigates the effect of exposure to ambient $PM_{2.5}$ in the month of birth. In an extension, we consider exposure at other ages. We average over three-month age ranges, from -9 to -7 months (first trimester of pregnancy) to 9 to 11 months (the last quarter of the first year of life). Average $PM_{2.5}$ in each age of exposure is used as the independent variable in a separate regression:

$$h_{ipdmy} = \beta \left( \frac{x_{dmy}^{+0} + x_{dmy}^{+1} + x_{dmy}^{+2}}{3} \right) + \boldsymbol{\alpha}_{pd} + \boldsymbol{\gamma}_{dm} + \boldsymbol{\delta}_{y} + X_{ipdmy}\boldsymbol{\theta} + \varepsilon_{ipdmy} \qquad (2)$$

where indices and fixed effects are as in regression equation (1), but the controls $X$ include only the age-in-months-by-sex indicators, and not the full set of extended controls. Coefficient estimates and 95% confidence intervals are presented in Figure 5.

**Robustness of shape of the concentration-response function.** The shape of the concentration-response function has been a focus of the air pollution literature, in light of its importance for policy responses (Pope et al. 2015). Although the prior literature has emphasized the possibility of diminishing marginal costs (such that the extra damages from extra exposure decline at higher levels of exposure), there is little well-identified evidence on exposure to $PM_{2.5}$ at levels as high as in India during the period studied.

Therefore, we perform three robustness checks in which we allow the shape of the concentration-response function to differ from the linear form in equation (1):

$$h_{ipdmy} = f(x_{dmy}) + \boldsymbol{\alpha}_{pd} + \boldsymbol{\gamma}_{dm} + \boldsymbol{\delta}_{y} + X_{ipdmy}\boldsymbol{\theta} + \varepsilon_{ipdmy}. \qquad (3)$$

First, in Table 2, we substitute in the natural log of $PM_{2.5}$ in one specification, and a linear spline at the median of $PM_{2.5}$ in another. Then, in Figure 6, we allow polynomial shapes of the concentration-response curve, of degree 1 through 5. Finally, in Figure 7, we implement a Box-Cox power transformation of the form $f(x) = x^{\lambda}$, for coefficients $\lambda$ in steps of 0.1 from 0.1 to 2.0. We implement each power transformation in a separate regression, and plot the resulting log-likelihoods. If likelihood is maximized near $\lambda = 1$, then this procedure would suggest that a linear concentration-response function fits these data.

**Acknowledgments.** SD acknowledges DST-FIST grant (SR/FST/ESII-016/2014) for upgradation of computing facility at IIT Delhi. MISR aerosol data are archived in NASA Langley Research Atmospheric science data Center.

**Code availability.** The codes used to arrive at the results depicted in this study can be available upon request to the corresponding author.

## References

1. Šrám, R. J., Binková, B., Dejmek, J. & Bobak, M. Ambient air pollution and pregnancy outcomes: A review of the literature. *Environ. Health Perspect.* **113,** 375–382 (2005).

2. Litchfield, I. J., Ayres, J. G., Jaakkola, J. J. K. & Mohammed, N. I. Is ambient air pollution associated with onset of sudden infant death syndrome: a case-crossover study in the UK. *BMJ Open* **8,** (2018).

3. Balakrishnan, K. *et al.* Exposures to fine particulate matter (PM2.5) and birthweight in a rural-urban, mother-child cohort in Tamil Nadu, India. *Environ. Res.* **161,** 524–531 (2018).

4. Payam, D. *et al.* Maternal Exposure to Particulate Air Pollution and Term Birth Weight: A Multi-Country Evaluation of Effect and Heterogeneity. *Environ. Health Perspect.* **121,** 267–373 (2013).

5. Mishra, V. & Retherford, R. D. Does biofuel smoke contribute to anaemia and stunting in early childhood? *Int. J. Epidemiol.* **36,** 117–129 (2007).

6. Balietti, A. & Datta, S. The impact of indoor solid fuel use on the stunting of Indian children. (2017).

7. Goyal, N. & Canning, D. Exposure to Ambient Fine Particulate Air Pollution in Utero as a Risk Factor for Child Stunting in Bangladesh. *International Journal of Environmental Research and Public Health* **15,** (2018).

8. Chowdhury, S. & Dey, S. Cause-specific premature death from ambient PM2.5 exposure in India: Estimate adjusted for baseline mortality. *Environ. Int.* **91,** (2016).

9. Dey, S. *et al.* Variability of outdoor fine particulate (PM2.5) concentration in the Indian Subcontinent: A remote sensing approach. *Remote Sens. Environ.* **127,** 153–161 (2012).

10. Coffey, D. & Spears, D. E. *Where India goes : abandoned toilets, stunted development and the costs of caste: Harper Collins*. (2017).

11. Coffey, D. & Spears, D. Child Height in India. *Economic and Political Weekly* 87–94 (2018).

12. Currie, J. & Vogl, T. Early-Life Health and Adult Circumstance in Developing Countries. (2013). doi:10.1146/annurev-economics-081412-103704

13. Rangel, M. & Vogl, T. Agriculture, Fire, and Infant Health ∗. *Natl. Bur. Econ. Res.* (2017).

14. Spears, D. Exposure to open defecation can account for the Indian enigma of child height. *J. Dev. Econ.* (2018). doi:https://doi.org/10.1016/j.jdeveco.2018.08.003

15. Gaudermann, W. *et al.* Association between Air Pollution and Lung Function Growth in Southern California Children. *Am. J. Respir. Crit. Care Med.* **162,** 1383–1390 (2000).

16. Bozzolli, C., Deaton, A. & Quitana-Domeque, C. Adult Height and Childhood Disease. *Demography* **46,** 647–669 (2009).

17. Setel, P. W. *et al.* A scandal of invisibility: making everyone count by counting everyone. *Lancet* **370,** 1569–1577 (2007).

18. Spears, D., Ghosh, A. & Cumming, O. Open Defecation and Childhood Stunting in India: An Ecological Analysis of New Data from 112 Districts. *PLoS One* **8,** e73784 (2013).

19. Chowdhury, S., Dey, S. & Smith, K. R. Ambient PM2.5 exposure and expected premature mortality to 2100 in India under climate change scenarios. *Nat. Commun.* **9,** 318 (2018).

20. Spears, D. Height and cognitive achievement among Indian children. *Econ. Hum. Biol.* **10,** 210–219 (2012).

21. van Donkelaar, A. *et al.* Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **118,** 847–855 (2010).

22. van Donkelaar, A., Martin, R. V., Brauer, M. & Boys, B. L. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ. Health Perspect.* **110,** 135–143 (2014).

23. Dey, S. & Di Girolamo, L. A climatology of aerosol optical and microphysical properties over the Indian subcontinent from 9 years (2000-2008) of Multiangle Imaging Spectroradiometer (MISR) data. *J. Geophys. Res. Atmos.* **115,** 1–22 (2010).

24. Victora, C. G., de Onis, M., Hallal, P. C., Blössner, M. & Shrimpton, R. Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions. *Pediatrics* **125,** e473 LP-e480 (2010).

25. Deaton, A. *The analysis of household surveys. A microeconometric approach to development policy : The World Bank. The analysis of household surveys: A microeconometric approach to development policy* (1997). doi:doi:10.1596/0-8018-5254-4

26. Darrouzet-Nardi, A. F. & Masters, W. A. Urbanization, market development and malnutrition in farm households: evidence from the Demographic and Health Surveys, 1986–2011. *Food Secur.* **7,** 521–533 (2015).

27. Geruso, M. & Spears, D. Neighborhood sanitation and infant mortality. *Am. Econ. J. Appl. Econ.* **10,** 125–162 (2018).

28. Barreca, A., Deschenes, O. & Guldi, M. Maybe Next Month? Temperature Shocks and Dynamic Adjustments in Birth Rates. *Demography* **55,** 1269–1293 (2018).

29. Buckles, K. S. & Hungerman, D. M. Season of Birth and Later Outcomes: Old Questions, New Answers. *Rev. Econ. Stat.* **95,** 711–724 (2013).

30. Larsen, A. F., Headey, D. & Masters, W. A. Misreporting Month of Birth: Implications for Nutrition Research. *Work. Pap.* 1–56 (2017).

31. Agarwal, N. *et al.* Month of birth and child height in 40 countries. *Econ. Lett.* **157,** 10–13 (2017).

32. Colin Cameron, A. & Miller, D. L. A Practitioner's Guide to Cluster-Robust Inference. *J. Hum. Resour.* **50,** 317–372 (2015).

33. Solon, G., Haider, S. J. & Wooldridge, J. What Are We Weighting For? *J. Hum. Resour.* **50,** 301–316 (2015).

**Author contributions.** D.S. and S.D. developed the idea, D.S. carried out the main analysis, and S.C. carried out the satellite data analysis to generate the exposure data. D.S. and S.D. wrote the paper with inputs from all the authors.

**Competing financial interests.** The authors declare they have no actual or potential competing financial interests.

**Table 1.** Summary statistics describing height-for-age sample from India's 2015-16 DHS.

| | | $PM_{2.5}$ district-month of birth quintiles | | | | |
|---|---|---|---|---|---|---|
| | full sample | 1 | 2 | 3 | 4 | 5 |
| $PM_{2.5}$ in birth month | 54.9 | 15.3 | 30.1 | 45.7 | 65.2 | 118.2 |
| height-for-age | -1.50 | -1.35 | -1.45 | -1.52 | -1.59 | -1.60 |
| age in months | 30.7 | 31.9 | 31.2 | 30.7 | 30.4 | 29.1 |
| girl | 0.48 | 0.49 | 0.48 | 0.48 | 0.47 | 0.48 |
| rural | 0.72 | 0.67 | 0.71 | 0.72 | 0.74 | 0.76 |
| mother's height (cm) | 151.7 | 152.2 | 151.9 | 151.6 | 151.4 | 151.2 |
| uses LPG | 0.33 | 0.42 | 0.36 | 0.33 | 0.28 | 0.25 |
| uses traditional fuel | 0.63 | 0.53 | 0.60 | 0.63 | 0.68 | 0.72 |
| open defecation | 0.47 | 0.40 | 0.47 | 0.49 | 0.51 | 0.49 |
| birth order | 2.18 | 1.97 | 2.09 | 2.18 | 2.30 | 2.38 |
| sibsize | 2.46 | 2.23 | 2.37 | 2.46 | 2.59 | 2.67 |
| in UP or Bihar | 0.31 | 0.09 | 0.18 | 0.29 | 0.41 | 0.57 |
| *n* (children under 5) | 218,152 | 52,947 | 43,942 | 40,831 | 40,551 | 39,881 |

*Note:* Each number, other than sample sizes in the bottom row, is a sample mean. Girl, rural, uses LPG, uses traditional fuel, open defecation and in Uttar Pradesh (UP) or Bihar are each indicators (1 or 0) for that property of the child or household. Sibsize is a measure of fertility: the number of children born to the child's mother at the time of the survey. Sample means and quintiles are computed with DHS sampling weights (which is why *n* is not constant across quintiles).

**Table 2.** Regression results: Child height-for-age *z*-score regressed on district-level $PM_{2.5}$ in month of birth, with fixed effects and covariate controls

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| $PM_{2.5} \div 100$ | -0.0546** | -0.0494* | -0.0488* | -0.0477* | -0.0502* | | -0.0352* | -0.0734* | -0.0286 |
| | (0.0201) | (0.0210) | (0.0207) | (0.0206) | (0.0216) | | (0.0178) | (0.0309) | (0.0555) |
| $PM_{2.5} \div 100$ 24 months earlier | | | | | -0.0107 | | | | |
| | | | | | (0.0209) | | | | |
| $\ln(PM_{2.5})$ | | | | | | -0.0171$^{\dagger}$ | | | |
| | | | | | | (0.00987) | | | |
| mother's height (cm) | | | 0.0470** | 0.0467** | 0.0467** | 0.0467** | | | |
| | | | (0.000932) | (0.000925) | (0.000925) | (0.000925) | | | |
| $PM_{2.5} \div 100$ above median spline | | | | | | | | | -0.0281 |
| | | | | | | | | | (0.0668) |
| age-in-months × sex | yes | yes | yes | yes | yes | yes | yes | yes | yes |
| district-month FEs | yes | yes | yes | yes | yes | yes | | | yes |
| year of birth FEs | yes | yes | yes | yes | yes | yes | yes | yes | yes |
| PSU FEs | | yes | yes | yes | yes | yes | yes | | yes |
| extended controls | | | yes | yes | yes | yes | | | |
| household fuel | | | | yes | yes | yes | | | |
| state-month FEs | | | | | | | yes | | |
| PSU-month FEs | | | | | | | | yes | |
| *n* (children under 5) | 218,152 | 217,285 | 216,745 | 216,745 | 216,745 | 216,745 | 217,286 | 115,586 | 217,285 |

*Note:* All columns present ordinary least squares fixed effects regressions with the child's height-for-age *z*-score as the dependent variable. FE = fixed effect; PSU = primary sampling unit (urban block or rural village). Standard errors clustered by 640 districts in parentheses. † $p < 0.10$; * $p < 0.05$; ** $p < 0.01$. In column 9, the spline variable is zero below the median $PM_{2.5}$ and is identical to $PM_{2.5}$ above the median. Sample sizes vary because some fixed effects categories lack within-category variation in the independent variable (resulting in that category being dropped), and because not all children's mothers' height was measured.

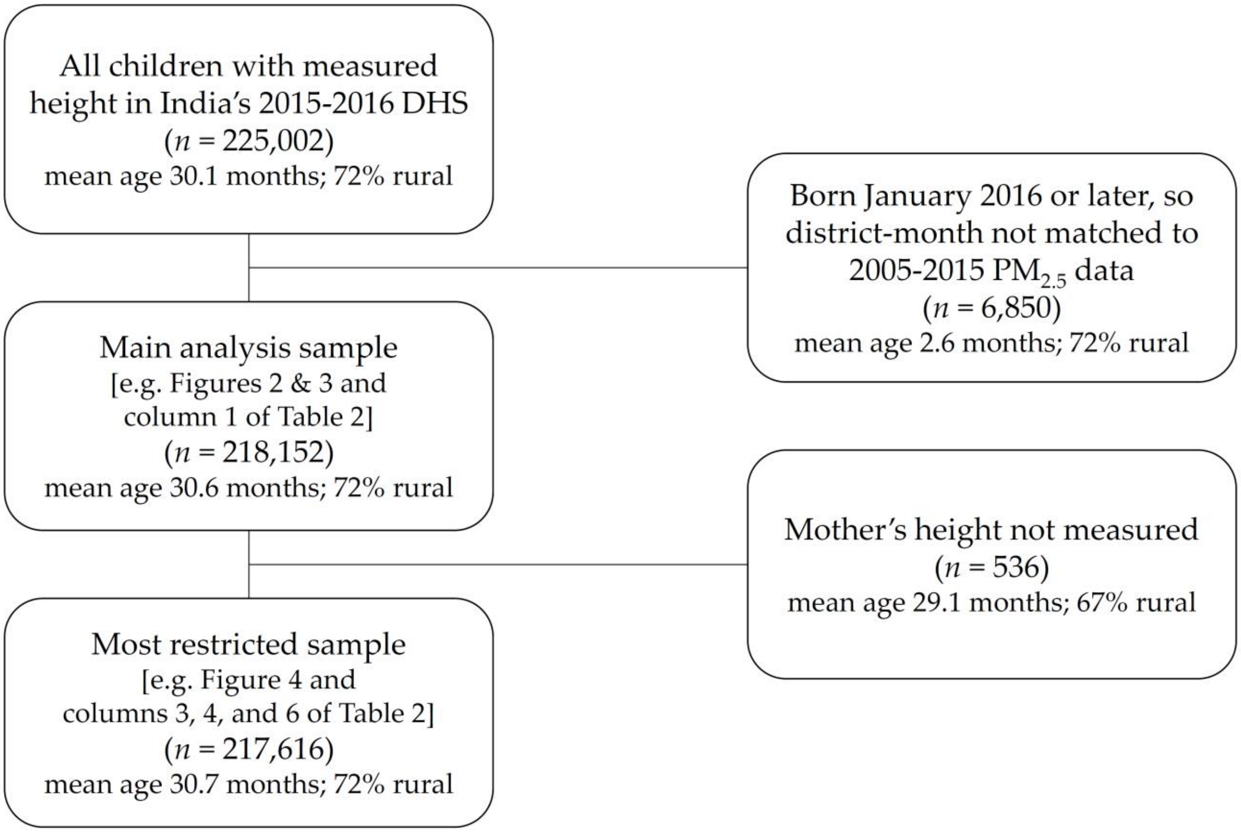


**Figure 1.** Study sample with excluded or missing observations. In Table 2, some samples are smaller than 217,616 because fixed effects regression ignores categories within which there is no variation in the independent variable.

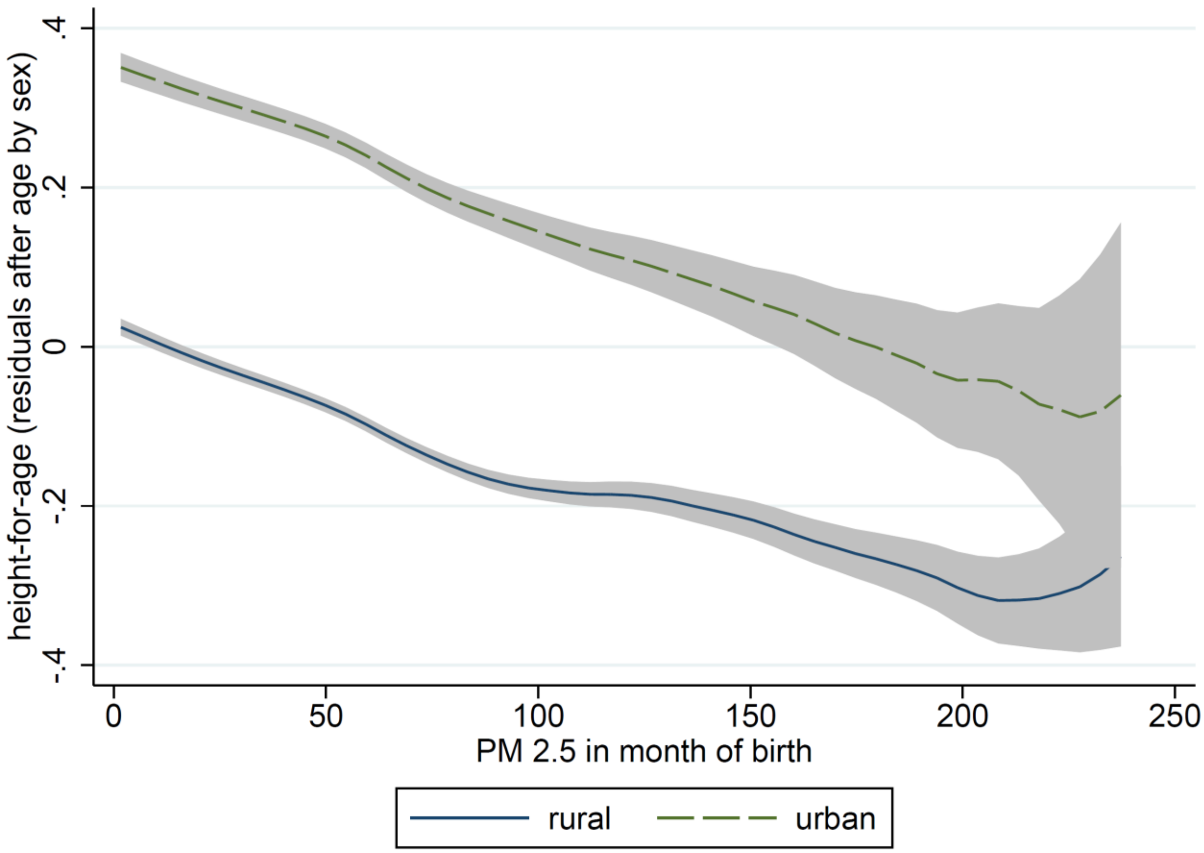


**Figure 2.** Non-parametric association between child height and exposure to $PM_{2.5}$ in the month of birth. Curves are kernel-weighted local regressions. The vertical axis is the residual of child height-for-age on 120 age-in-months by sex indicators.

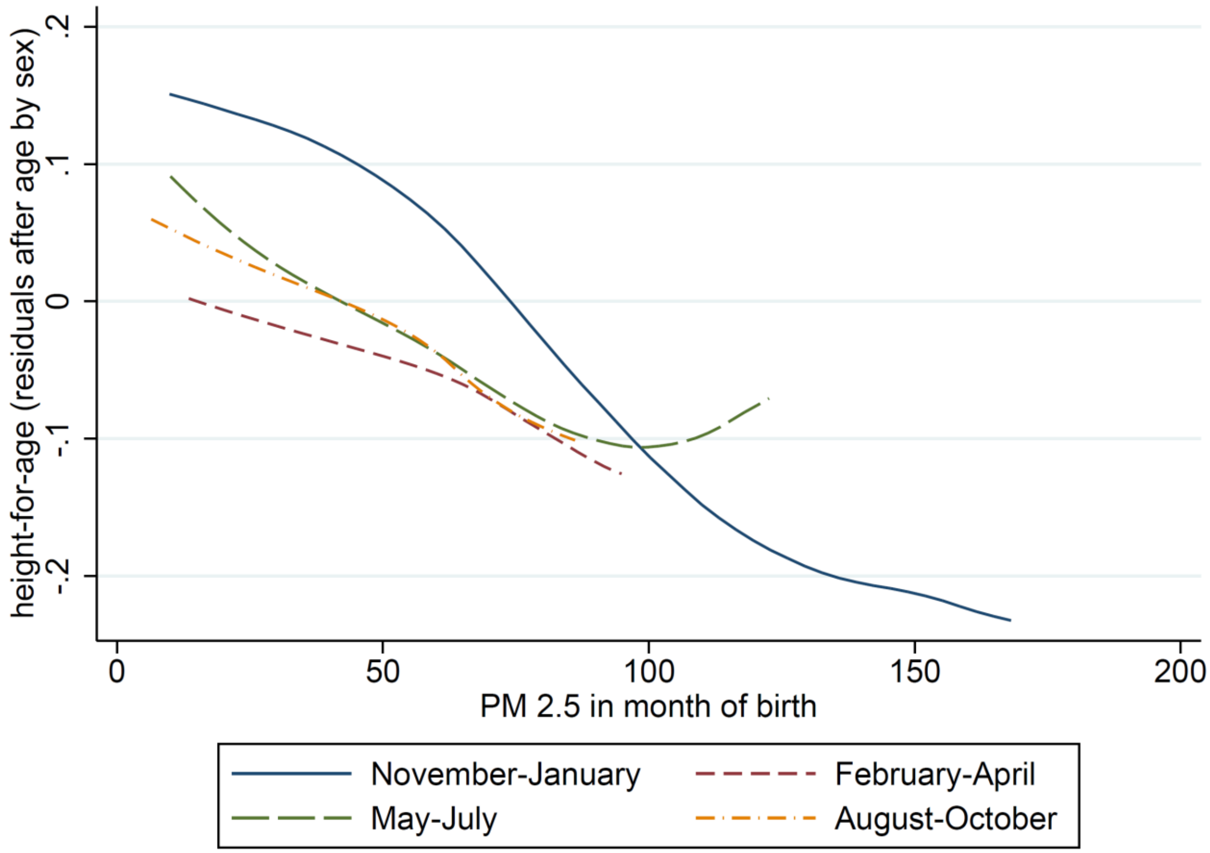


**Figure 3.** Child height and exposure to $PM_{2.5}$ in the month of birth, by season of birth. Curves are non-parametric kernel-weighted local regressions. The vertical axis is the residual of child height-for-age on 120 age-in-months by sex indicators.

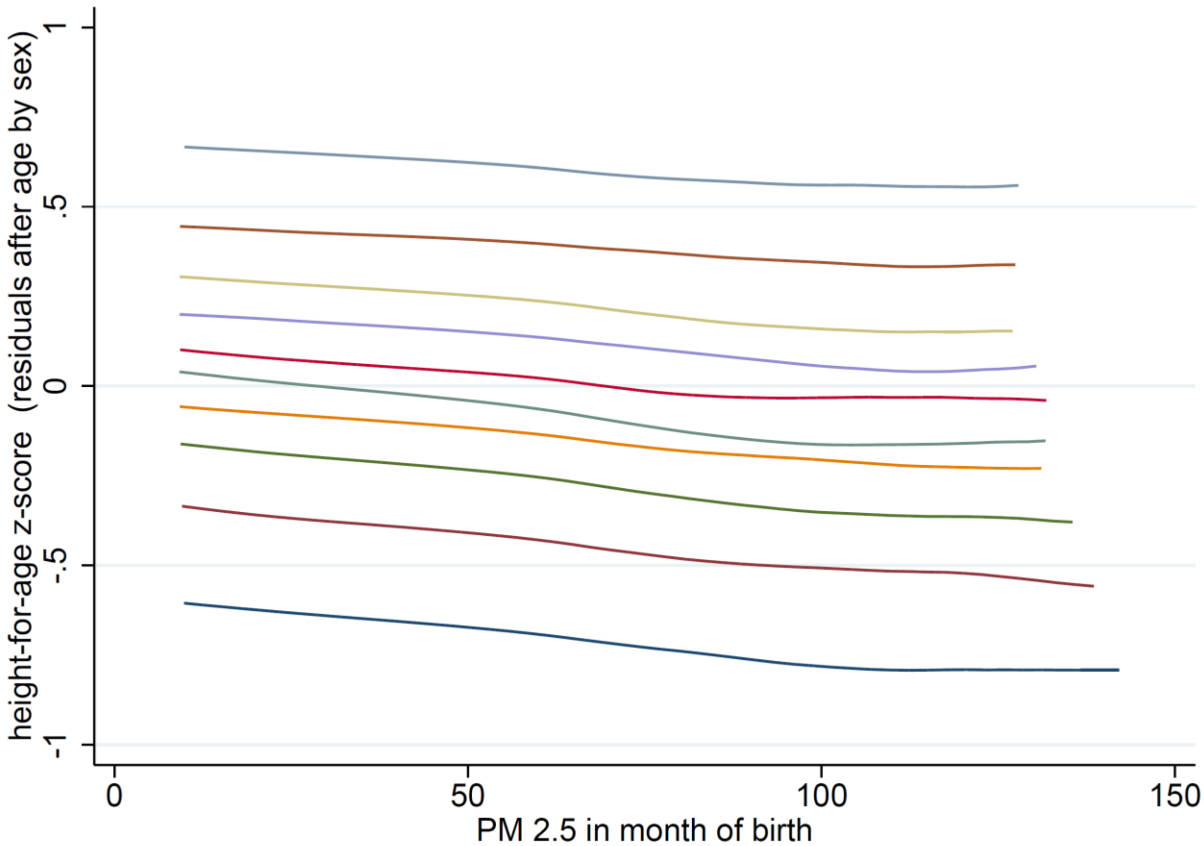


**Figure 4.** Child height and exposure to $PM_{2.5}$ in the month of birth, by decile of mother's height. Curves are non-parametric kernel-weighted local regressions. Each separate curve includes only children born to mothers in one of ten height deciles. The vertical axis is the residual of child height-for-age on 120 age-in-months by sex indicators.

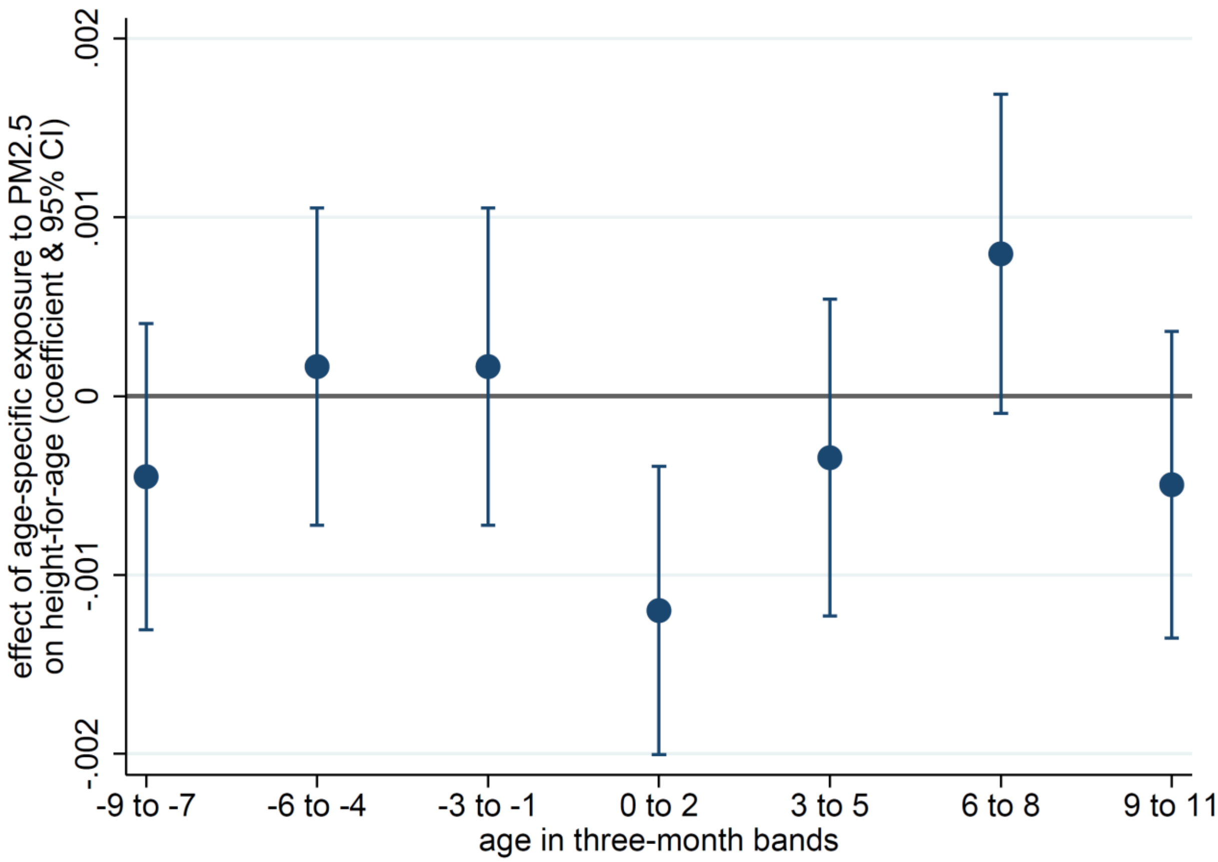


**Figure 5.** Effects of $PM_{2.5}$ exposure at various ages. Each dot is a coefficient and each range is a 95% confidence interval from a separate fixed effects regression of child height-for-age on the average exposure to $PM_{2.5}$ in the months, relative to birth, specified along the horizontal axis.

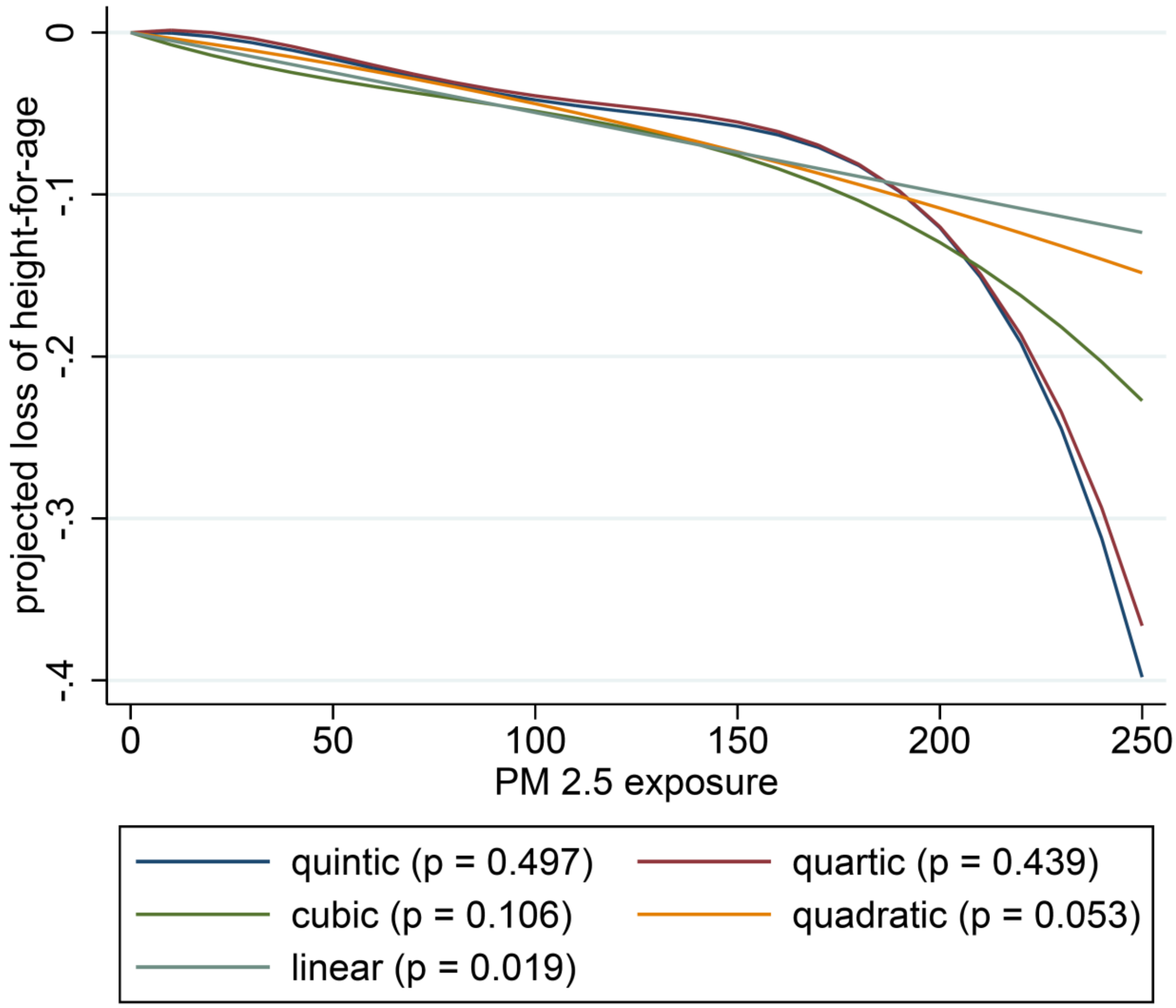


**Figure 6.** Projected effects of $PM_{2.5}$ of child height-for-age, at increasing non-linearity. Each curve is the projected effect from a separate fixed effects regression where $PM_{2.5}$ in the month of birth is specified as a polynomial of degree 1 through 5. *p*-values report joint *F* tests that all $PM_{2.5}$ terms are zero.

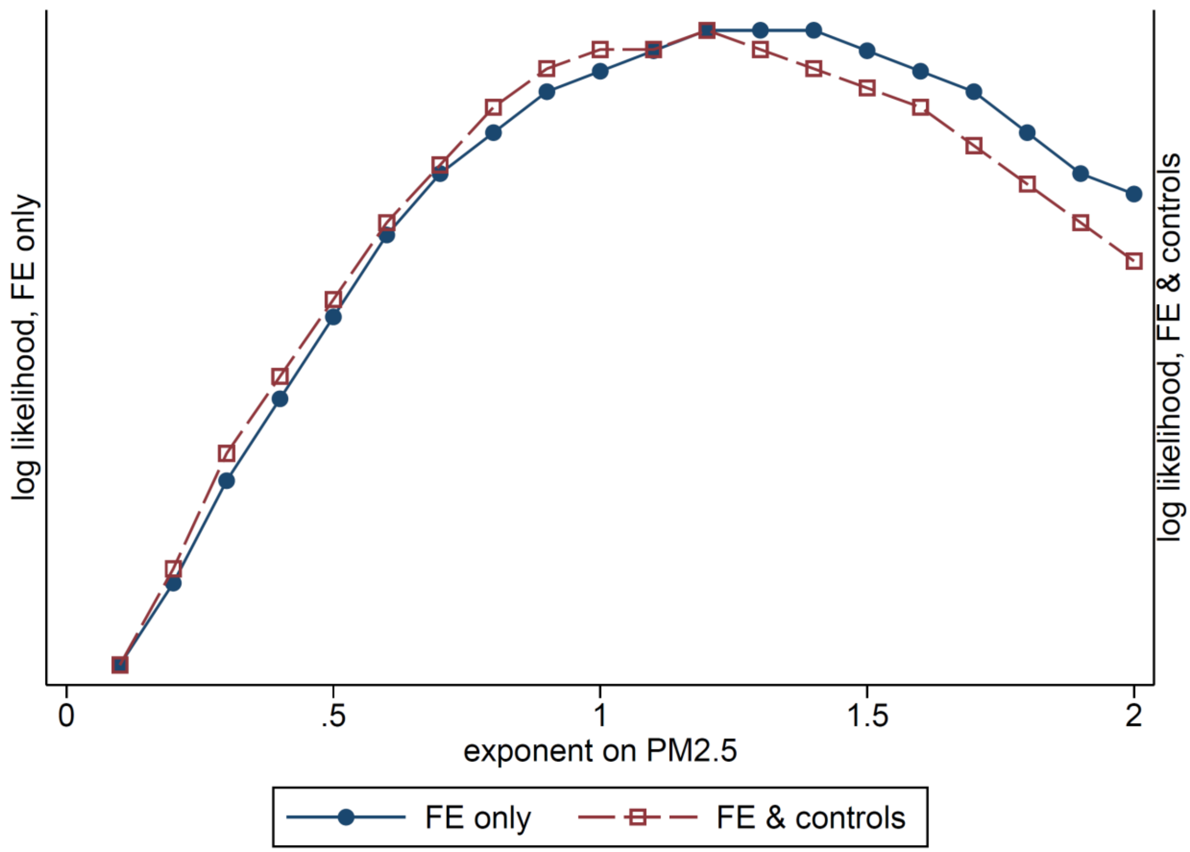


**Figure 7.** Box-Cox transformation of $PM_{2.5}$ in month of birth: Log likelihood. Each point plots the log likelihood of a separate fixed effects regression of $PM_{2.5}$ transformed according to the coefficient on the horizontal axis.