Health Effects of Sustained Exposure to Fine Particulate Matter: Evidence from India [a]

Yashaswini Saraswat [1,2] & Sangeeta Bansal [1]

Keywords: Fine particulate matter, sustained exposure, India, life expectancy, early life mortality, regression discontinuity.

---

[a] **JEL Classification:** I12, I15, Q51, Q53, Q56

[1] Centre for International Trade and Development, Jawaharlal Nehru University, Delhi-67
[2] Corresponding Author, E-mail: yashaswini.saraswat@gmail.com, Ph: +91 8800240084

ABSTRACT

Owing to their peculiar topography and location, the Indo-Gangetic Plains belong to the most polluted regions of the world; nine out of ten most polluted cities in the world lie here. The valley traps the particulate matter generated in the region and is hence exposed to two to four times higher fine particulate matter [particulate matter smaller than 2.5 μg/m^3 ($PM_{2.5}$)] than rest of the country. This study utilises the "valley effect" as a basis for a natural experiment. An exogenous threshold is drawn from geographical literature and is based on the lower boundary of the Indo-Gangetic Plains. Results obtained by using a regression discontinuity design in this study provide first causal estimates on the health impact of long-term exposure to $PM_{2.5}$ by using data from India. $PM_{2.5}$ exposure is found to be 49% higher and life expectancy is 2.6 years lower in the Plains relative to other districts in the sample. Early life mortality is found to be positively and significantly affected by sustained exposure to $PM_{2.5}$. It is also found that life expectancy at birth reduces by 1.2 years due to additional 10 μg/m^3 of $PM_{2.5}$ exposure, ceteris paribus. Around 5.2 years of life can be saved in the Indo-Gangetic Plains if the national standard for $PM_{2.5}$ is met in the region. The life years saved rise to 8.8 years when the WHO standard is met. India can raise life expectancy by 1.7 and 5.3 years if the national and the WHO standards for $PM_{2.5}$ are met, respectively.

HIGHLIGHTS

- Regression discontinuity design based on Indo-Gangetic Plains.
- Fall in life expectancy by 1.2 years due to additional 10 μg/m^3 of $PM_{2.5}$ exposure.
- Estimated gains in life expectancy of 5.3 years on meeting the WHO standards in India.

INTRODUCTION

Air pollution has reached dangerously high levels in India. More than 99% of the Indian population is exposed to $PM_{2.5}$ concentration levels exceeding the World Health Organization standard of 10 $\mu g/m^3$ (Apte and Pant, 2019). Higher incidence of cardiorespiratory diseases like heart stroke, lung cancer, asthma has been observed in heavily polluted regions in the world (WHO, 2016). Studies have also found early life mortality to rise and life expectancy to fall as a result of exposure to $PM_{2.5}$ (Pope et al., 2002; Tanaka, 2015; Heft-Neal et al., 2018). Ambient air pollution hinders the natural growth process in the early life stages (Zhang et al., 2018; Lavigne et al., 2018) and may even lead to a modification in the DNA of the population (Carre et al., 2017).

Several studies find a robust relationship between particulate matter and health in different countries. Chay and Greenstone (2003) have utilised a regression discontinuity design (RDD) to study the impact of total suspended particulate (TSP) matter on infant mortality rate (IMR) for US counties during 1971-72. They find a fall in infant mortality rate by 0.5% due to a 1% decline in TSP concentration. Arceo et al. (2016) find a rise in IMR by 8.8% due to a rise in PM_{10} by 10 $\mu g/m^3$ in Mexico city. Heft-Neal et al. (2018) find an IMR effect of 9.2% due to a rise in $PM_{2.5}$ by 10 $\mu g/m^3$ in Africa. Greenstone and Hanna (2014) conducted an analysis of the impact of environmental regulations targeting air and water pollution on infant mortality rate in India. They find that though the regulations improved air quality, it did not have a statistically significant impact on infant mortality rate. Water policies, however, did not lead to any improvement in water quality. In a study conducted in China, Tanaka (2015) finds a significant decline of 20% in infant mortality rate due to the first large-scale environmental regulations in China.

Many studies report that a reduction in long-term exposure to particulate matter improves life expectancy. Pope et al. (2002) find a gain of 0.7 years due to 10 $\mu g/m^3$ fall in the long term exposure to $PM_{2.5}$ in the United States. Chen et al. (2013) and Ebenstein et al. (2017) find gains in life expectancy to be 0.3 years and 0.64 years in China due to 10 $\mu g/m^3$ decline in the long term exposure to TSP and PM_{10}, respectively. Such studies continue to hold immense significance for environmental policy making in countries like India where the emission standards remain too high and a systematic tackling of the air pollution problem is urgently required. However, no such long-term exposure study on $PM_{2.5}$ that utilises data from Indian population exists.

The objective of this study is to estimate the impact of long-term exposure to $PM_{2.5}$ on health outcomes in India. More specifically, we examine the impact of sustained exposure to $PM_{2.5}$ on early life mortality and life expectancy at birth in India. For early life mortality rates, we consider neo-natal mortality rate, infant mortality rate and under-five mortality rate. Evidence on such effect is important because recent scientific studies have found that the small size of $PM_{2.5}$ enables it to enter the blood stream and placenta of pregnant women, thereby affecting the health status of their offspring (Zhang et al., 2018; Lavigne et al., 2018). Such evidence will facilitate better understanding of the benefits of reducing the concentration levels of $PM_{2.5}$ in the country. The results from the study will facilitate evidence based environmental policy making in India and other countries that face similar levels of air pollution.

To investigate the causal effect of air pollution on health, we identify a physical boundary from the geographical literature that serves as an exogenous threshold to facilitate a regression discontinuity design (RDD). According to the Plate Tectonic Theory, collision of the Indian subcontinent with the Eurasian continent led to the origin of the Himalayan Ranges and a deep depression on their south (Burrard, 1915; Aitchison and Davis, 2007). This depression, known as the Indo-Gangetic Plains (IGP), is one of the most populated as well as polluted regions of the world. The peculiar topography of the region being sunken and landlocked by the Himalayas on the North and the Central Highlands on its South restricts the wind passage thereby making displacement of the particulate matter generated in the region difficult (Guttikunda and Gurjar, 2012). Districts lying south of the Plains do not suffer from this unfavorable "valley effect" and hence are exposed to much lower pollution levels.

For the study, we consider districts in the Indo-Gangetic Plains and districts that lie below the plains, but within a distance of approximately five degree latitude. The lower boundary of the Indo-Gangetic Plains (Fig. 1) exogenously divides this sample into treatment and control group such that the districts in the Indo-Gangetic Plains form the treatment group while the other districts form the control group. The districts to the north and the south of the boundary are similar in several ways and some of them even lie in the same state. Differences in ecology, degree of urbanization, and socioeconomic variables may influence health indicators such as life expectancy and early life mortality. We check for the validity of a regression discontinuity design by using predicted life expectancy as a proxy variable for the health impact of these factors (Chen et al., 2013; Ebenstein et al. 2017). We find that predicted life expectancy moves smoothly across the boundary as it is insignificantly different at the boundary. This helps in examining a causal relationship between human health and air pollution. Moreover, we control for variables like income, literacy rate, share of rural households, share of minority population, share of households with access to clean drinking water and clean cooking fuel in our regression equations.

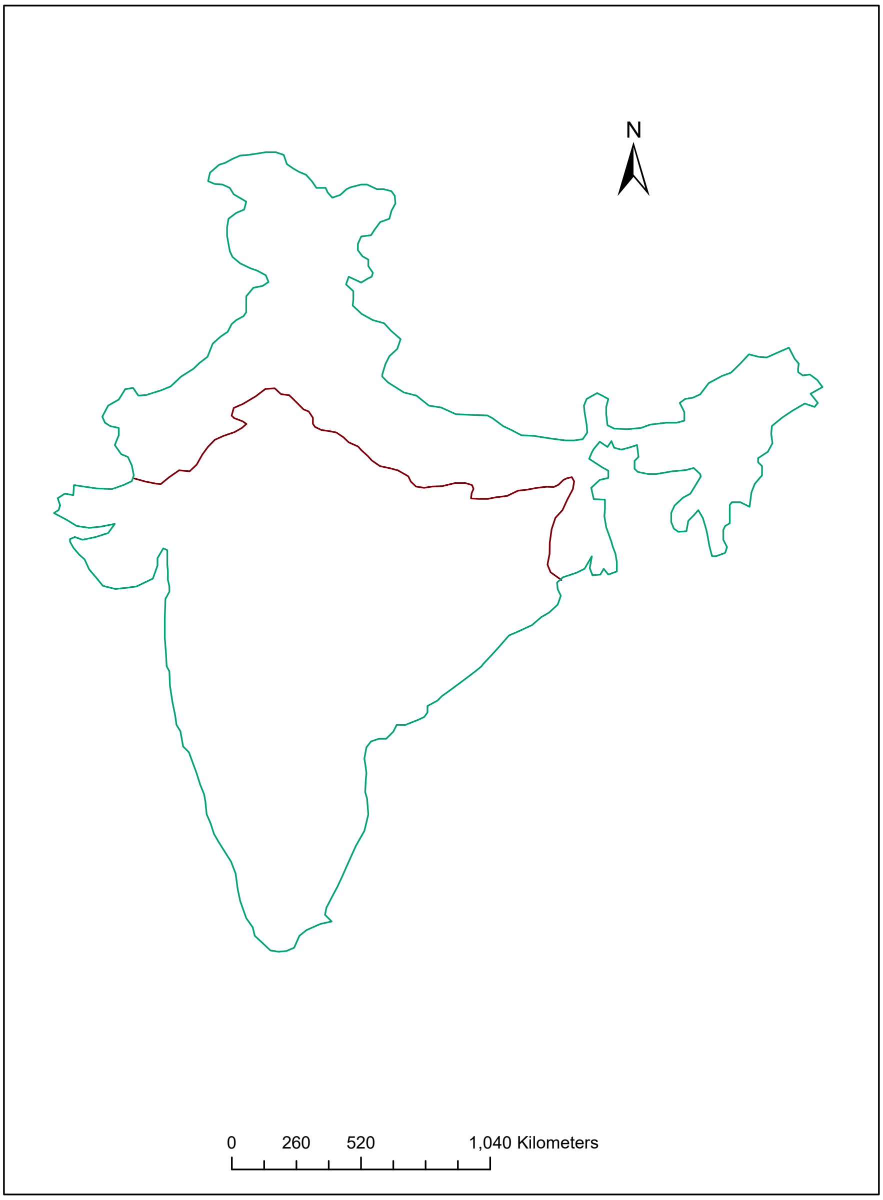

Fig. 1. Lower boundary of the Indo-Gangetic Plains depicted in brown, used as the threshold for conducting Regression Discontinuity Analysis

Source: Plotted using ArcGIS

Our findings are as follows. The unfavorable location and topography of the Indo-Gangetic Plains has severely affected the pollution concentrations in the region and the health of its inhabitants. The $PM_{2.5}$ exposure is about 23 $\mu g/m^3$ or 49% higher, and life expectancy is 2.6 years lower in the region as compared to the control group. The early life mortality rates are also significantly higher in the region suggesting that when individuals are exposed to high levels of pollution for a long period of time, the probability of survival of their future offspring may reduce considerably. Further, life expectancy at birth reduces by 1.2 years due to additional 10 $\mu g/m^3$ of $PM_{2.5}$ exposure, ceteris paribus. Hence, India can raise life expectancy on average by 1.7 and 5.3 years if the national and WHO standards for $PM_{2.5}$ are met, respectively.

This study addresses several gaps in the literature assessing the health impact of air pollution. First, the study estimates the impact of sustained exposure to $PM_{2.5}$ on life expectancy and early life mortality in India by utilizing differences in long run exposure to particulate matter in the Indo-Gangetic Plains (IGP) and the districts below it. The studies from India have estimated health effects of short-term exposure to particulate matter (Guttikunda and Goel, 2013; Chowdhury and Dey, 2016). These studies often underestimate the loss in life expectancy since they only capture deaths of the vulnerable population such as old and sick that are accelerated due to a sudden rise in air pollution (Lvovsky, 1998).

Second, this is the first long-term exposure study that utilizes data from India. Scarcity of reliable data in India has limited the number of long-term exposure studies for India. Greenstone et al. (2015) have utilized the estimates from the Chinese study by Chen et al. (2013) to extrapolate the life expectancy gains in India of 3.2 years due to a reduction in $PM_{2.5}$ levels to the annual Indian National Ambient Air Quality Standard of 40 $\mu g/m^3$. Similarly, Greenstone and Fan (2018) have used estimates from the study by Ebenstein et al. (2017) to find life expectancy gains of 1.8 years on achieving the Indian standards for $PM_{2.5}$. There is a substantial difference in the reported life expectancy gains in the above studies. A major limitation of these two studies is that the life expectancy gains due to reduction in $PM_{2.5}$ in India are computed by using coefficients that were estimated for China . Further, the studies by Chen et al. (2013) and Ebenstein et al. (2017) utilize data on TSP and PM_{10} exposure, respectively, to estimate their impact on mortality and life expectancy in China. Greenstone et al. (2015) and Greenstone and Fan (2018), then, use these estimates to extrapolate the health impact of $PM_{2.5}$ exposure in India by utilizing conversion ratios.

Arguably, these extrapolated estimates may not correctly reflect the health effects of improvements in air quality in India. The divergence may further be exacerbated due to several socioeconomic and behavioral differences between India and China. In such a scenario, it is important to estimate the impact of improvement in air quality on the health status of the Indian population using data from within India.

A novelty of the study is to use the lower boundary of IGP as an exogenous threshold for the regression discontinuity design to investigate a causal relationship between health and air pollution thereby advancing the literature from observational studies. Fourth, most studies have dealt with the mega-cities in India (Cropper et al., 1997; Shah and Nagpal, 1997a; Kandlikar and Ramachandran, 2000; Nema and Goyal, 2010; WHO, 2018). Although nine of the ten most polluted cities in the world lie in the IGP (Guttikunda and Jawahar, 2012), there has been limited literature covering the entire region. The Indo-Gangetic Plains are the main focus of this study. Fifth, the study fills the deficit in the literature on the health impact of $PM_{2.5}$ exposure in India. $PM_{2.5}$ is much more harmful to health than other pollutants due to its easy penetration in lungs (since the diameter of $PM_{2.5}$ is below 2.5 μm).

METHODS

We adopt three main approaches in our analysis, namely, the conventional Ordinary Least Squares method (OLS), the Regression Discontinuity Design (RDD) and the Two-Stage Least Squares method (2SLS). Under the OLS, the four health outcomes are regressed on $PM_{2.5}$ exposure as shown in the equation below:

$$Y_j = \alpha_0 + \alpha_1 PM_j + X_j \phi + e_j \tag{1}$$

where j represents a district in the sample. Y_j represents health outcome in district j, i.e., either life expectancy at birth or an early-life mortality rate. PM_j is the level of fine particulate matter exposure in district j over a period of 17 years, from 1998 to 2014. X_j is a vector of demographic covariates in a district which are likely to affect the mortality rate or expected life years. e_j is the random error which is assumed to be distributed normally with mean zero.

The ideal way to establish causality would be to compare the health outcomes of the same individuals with and without exposure to pollution. That is, the effect of a treatment must be compared with a counterfactual which is identical to the treatment group in every way but has not received the treatment. This means that one is interested in understanding the difference between $Y_i(1)$ and $Y_i(0)$ where $Y_i(k)$ is the outcome variable (i.e., health outcome) with $k = 1$ if individual i received the treatment (high exposure to fine particulate matter) and 0, otherwise. However, the pair $Y_i(1)$ and $Y_i(0)$ is never observed simultaneously since individual i is either exposed or not exposed to the treatment.
This lack of counterfactual may lead to omitted variable bias and hence to the existence of endogeneity.[1] In addition, data on fine particulate matter may have some measurement error as it has been captured through satellite. The satellite-driven data on pollution is unable to collect information on days that are too cloudy; these days could be more polluted relative to clear sunny days (Greenstone and Fan, 2018). The intensity of this problem may be reduced considerably with the use of averages.

The regression discontinuity design (RDD) was introduced by Thistlethwaite and Campbell (Thistlethwaite and Campbell, 1960) and has been extensively used in the literature to infer causality in diverse fields. In an RDD, whether an individual receives the treatment or not is completely or partially determined by an assignment variable (latitudinal difference in our analysis), which may take value on either side of an exogenously defined threshold. The regression discontinuity design then estimates the local average treatment effects by comparing the districts just above and just below the IGP boundary. If any discontinuity is observed in the conditional distribution of the outcome variable at the threshold, then it may be inferred as an evidence for a causal relationship between the treatment and the outcome (Imbens and Wooldridge, 2007). A necessary identifying assumption is that there should not be a discontinuous jump in the unobserved determinants of the outcome variable. The study utilizes the discontinuity in $PM_{2.5}$ at the IGP boundary that has emerged as a result of the peculiar topography and location of the Great Plains.

[1] The presence of endogeneity is verified in the model by using the Durbin-Wu-Hausman test.

The following equations are estimated to test whether the valley effect of the IGP causes a discontinuous change in the particulate exposure and health outcomes:

$$PM_j = \beta_0 + \beta_1 N_j + \beta_2 f(L_j) + X_j k + v_j \tag{2}$$

$$Y_j = \gamma_0 + \gamma_1 N_j + \gamma_2 f(L_j) + X_j \phi + u_j \tag{3}$$

where j refers to a district in the sample, PM_j is the average annual exposure to $PM_{2.5}$ in district j over the period 1998-2014. Y_j is the health outcome in district j. The health outcomes considered are neonatal mortality rate (NNMR), infant mortality rate (IMR), under-five mortality rate (U5MR), and life expectancy at birth (LEB). N_j is the dummy variable which takes value one if a district lies in the IGP, and zero, otherwise. $f(L_j)$ is a function of the degrees north of the IGP boundary. Separate regressions using $f(L_j)$ as a linear function and as a quadratic function are conducted. X_j is a vector of other covariates that may affect health outcomes such as income, literacy rate, access to clean drinking water, access to clean cooking fuel, share of rural households and minority share in the population in a district. β_1 and γ_1 are the coefficients of interest in the above equations, measuring the valley effect of the Indo-Gangetic Plains on exposure to fine particulate matter and the health outcomes, respectively.

To find the impact of $PM_{2.5}$ exposure on early life mortality and life expectancy at birth we use two-stage least squares method where $PM_{2.5}$ exposure is instrumented by dummy variable N_j. The exclusion restrictions for the instrument are satisfied. Equation (2) serves as the first stage equation, and the predicted values of PM_j are then used in the second stage, as shown in equation (4) below.

$$Y_j = \delta_0 + \delta_1 \widehat{PM}_j + \delta_2 f(L_j) + X_j \phi + e_j \tag{4}$$

where $\widehat{PM}_j$ denotes the fitted values obtained by estimating equation. The coefficient δ_1 estimates the average treatment effect on health outcomes in our analysis.

DATA SOURCES

Data on early life mortality rates viz., neonatal mortality rate (NNMR), infant mortality rate (IMR) and under-five mortality rate (U5MR) are computed from the National Family Health Survey (NFHS), Round 4, 2015-16. In the absence of data on life expectancy and age-specific death rates at district level, we use the United Nations method to compute life expectancy at birth (LEB). This method has been used by the Population Branch of the United Nations, Department of Social Affairs to compute life expectancy using only infant mortality rate in countries lacking data on vital statistics (Kesarwani, 2015). The underlying assumption here is that each district follows the same fertility and mortality patterns as the state it belongs to. In this method, first, LEB is estimated as a function of IMR separately for each state using the following linear regression equation.

$$Ln(LEB_s) = a + b * IMR_s \tag{5}$$

where subscript s denotes variable at the state level. The values of a and b thus estimated are used to compute LEB at district level by using the formula below.

$$LEB_d = exp(\hat{a} + \hat{b} * IMR_d) \tag{6}$$

where subscript d denotes variable at the district level, $\hat{a}$ and $\hat{b}$ denote the estimated values of a and b from equation (5). The methodology uses two data sources. The state-level data on life expectancy and infant mortality rate for a period from 1995-99 to 2012-16 are taken from the Sample Registration System (SRS) Abridged Life Table and the SRS Bulletin Reports, respectively. The data on infant mortality rate have been computed for 234 districts from NFHS (2015-16) using DHS guidelines.

District-level data on exposure to $PM_{2.5}$ was obtained from Donkelaar et al. (2016) for 233 districts in the sample from 1998 to 2014. "$PM_{2.5}$ exposure" for a given year is defined as the average of the $PM_{2.5}$ readings in all the previous years. For instance, the $PM_{2.5}$ exposure relevant to the year 2000 is calculated as the average of the $PM_{2.5}$ concentrations in the years 1998, 1999 and 2000. Subsequently, $PM_{2.5}$ exposure is averaged across all years for each district to arrive at the average $PM_{2.5}$ exposure. This is used as a measure for the sustained level of $PM_{2.5}$ exposure in each district. The data includes districts from the states of Bihar, Haryana, Punjab, Rajasthan, Madhya Pradesh, Uttar Pradesh and West Bengal and the union territory of Delhi. Jharkhand and Chhattisgarh have been excluded since they are newly formed states and the districts have been reorganized over time, which makes district level pollution data not comparable. The coordinates for the lower boundary of the IGP have been taken from the *Geomorphic Atlas of Indo-Gangetic Plains* (Hecht and Sinha, 2003) created by the University of Technology, Dresden, Germany and the Indian Institute of Technology, Kanpur, India, with the help of ArcGIS software. The latitudinal difference between each district and the boundary is then estimated. Data on the demographic determinants of health are taken from Census 2011, NSSO-Round 68 and NFHS-4.

RESULTS

Validity of Regression Discontinuity Design

Table 1 presents the summary statistics of the key variables in our analysis. Columns (1) and (2) report the mean values of the variables in the districts lying north and south of the IGP boundary, respectively. Column (3) reports the difference in the mean values of the variables found in columns (1) and (2). Column (4) shows the difference in means of the variables once they are adjusted for a quadratic function of the difference in the degrees of latitude. This serves as a test for discontinuity at the IGP boundary. Rejection of the null hypothesis for a variable implies that there is a significant difference in the mean values of the variable on the two sides of the boundary.

$PM_{2.5}$ exposure is observed to be significantly higher in the districts lying above the IGP boundary in Table 1. This implies that the northern districts lying in the IGP are much more polluted (about 1.6 times) than the districts lying below the IGP boundary. On an average, the actual life expectancy at birth is lower by 1.77 years in the northern districts lying in the IGP relative to the southern districts in the sample. Checking if the observed determinants of health differ significantly across the boundary, we find that while the difference in income, literacy rate, and share of rural households is insignificant, the share of minority population is significantly lower, access to clean drinking water and access to clean cooking fuel is significantly higher in the treatment group. The direction of these latter differences is likely to lead to improvements in health outcomes in the treatment group, thus, unlikely to weaken the basis of our analysis. We further test the validity of the RDD in the IGP setting by examining the predicted life expectancy in the two groups. Predicted life expectancy is estimated as the fitted value from an OLS regression of life expectancy at birth on demographic and socioeconomic determinants of health except $PM_{2.5}$. The difference in the predicted life expectancy is found to be insignificant at the boundary.

Table 1. Summary Statistics

Variable	North (1)	South (2)	Difference in means (3)	Adjusted difference in means (4)
I. Air pollution exposure				
$PM_{2.5}$, μg/m^3	74.52	47.92	26.60***	22.49***
II. Demographic characteristics				
Literacy rate	0.68	0.66	0.02	0.01
Share, rural households	0.77	0.79	-0.02	-0.03
Share, treated tap water	0.20	0.11	0.08***	0.10***
Share, Clean Cooking Fuel	0.33	0.25	0.08***	0.11***
Share, Minority	0.76	0.83	-0.07***	-0.06***
Ln(Income)	12.04	11.98	0.06	0.06
Predicted Life Expectancy	67.37	67.40	0.03	0.16
Actual Life expectancy	66.69	68.46	-1.77***	-1.82***

Note: n = 234. The results in column (4) are adjusted for a quadratic in degrees of latitude north of the Indo-Gangetic Plains boundary. Predicted life expectancy is calculated by OLS using all the demographic covariates shown.

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Fig. 2 illustrates the fitted values of $PM_{2.5}$ exposure in districts against the degrees north of the IGP boundary. The fitted line is obtained by estimating equation (2) excluding the covariates' vector X. There is a striking discontinuous jump in the $PM_{2.5}$ exposure at the boundary implying that geography of the IGP has caused extremely high pollution levels in districts lying in the Indo-Gangetic Plain relative to those districts which are close to the boundary but lie on its other side. There is a discontinuous jump of 22.5 μg/m^3 of $PM_{2.5}$ (around 47%) at the boundary.

Fig. 3 illustrates the fitted values of life expectancy at birth obtained by estimating equation (3) while excluding the covariates' vector X. A discontinuous fall of approximately 1.8 years is seen at the boundary. This suggests that there is a significant difference in life expectancy at birth in districts lying just above and below the IGP. Figs. 2 and 3 collectively suggest causality between fine particulate matter exposure and life expectancy since the discontinuous rise in $PM_{2.5}$ and the discontinuous fall in life expectancy occur at precisely the same location.

Fig. 4 presents the graphical analysis of the test for internal validity of the RDD. The fitted line is obtained by regressing predicted life expectancy at birth on a quadratic function in latitude. The included variables explain about 13% of the variation in life expectancy. However, there is an insignificant difference in the predicted life expectancy at the IGP boundary. This implies that predicted life expectancy (excluding the impact of $PM_{2.5}$) is equal in the districts lying just north and south of the IGP boundary. The insignificant difference in predicted life expectancy provides support to the validity of our RDD, and suggests that the demographic and socioeconomic variables are unable to explain the abrupt fall in life expectancy in districts lying just above the IGP boundary as shown in Fig. 3.

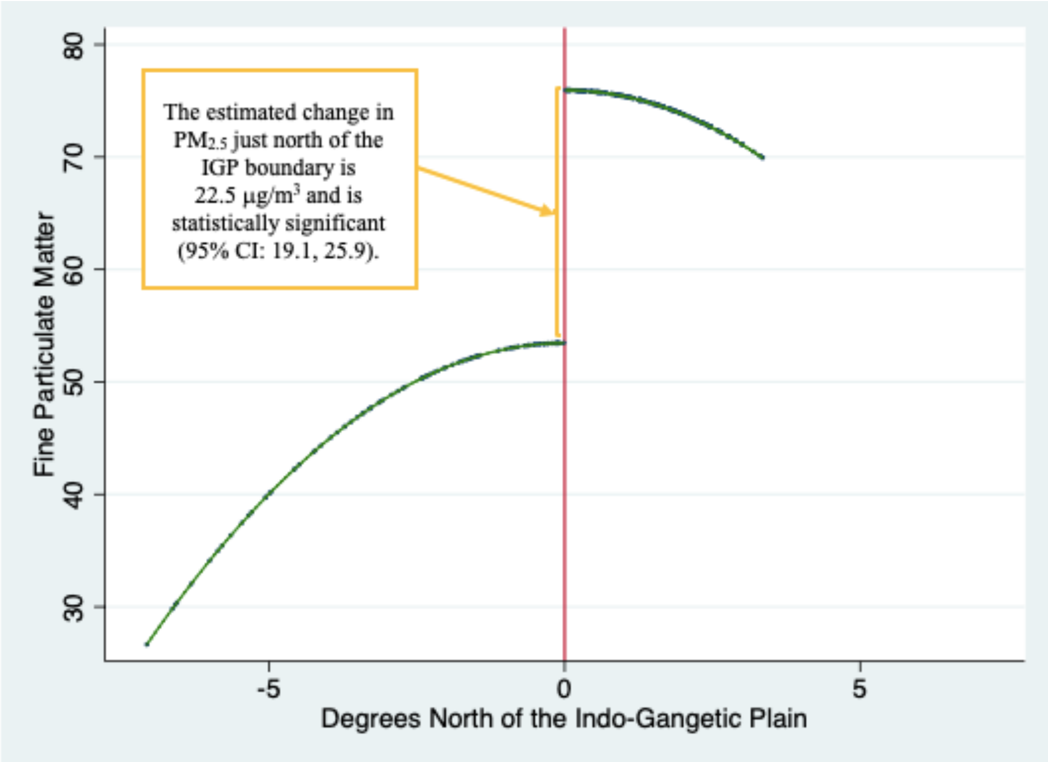

Fig. 2. Fitted values of $PM_{2.5}$ exposure across the IGP boundary

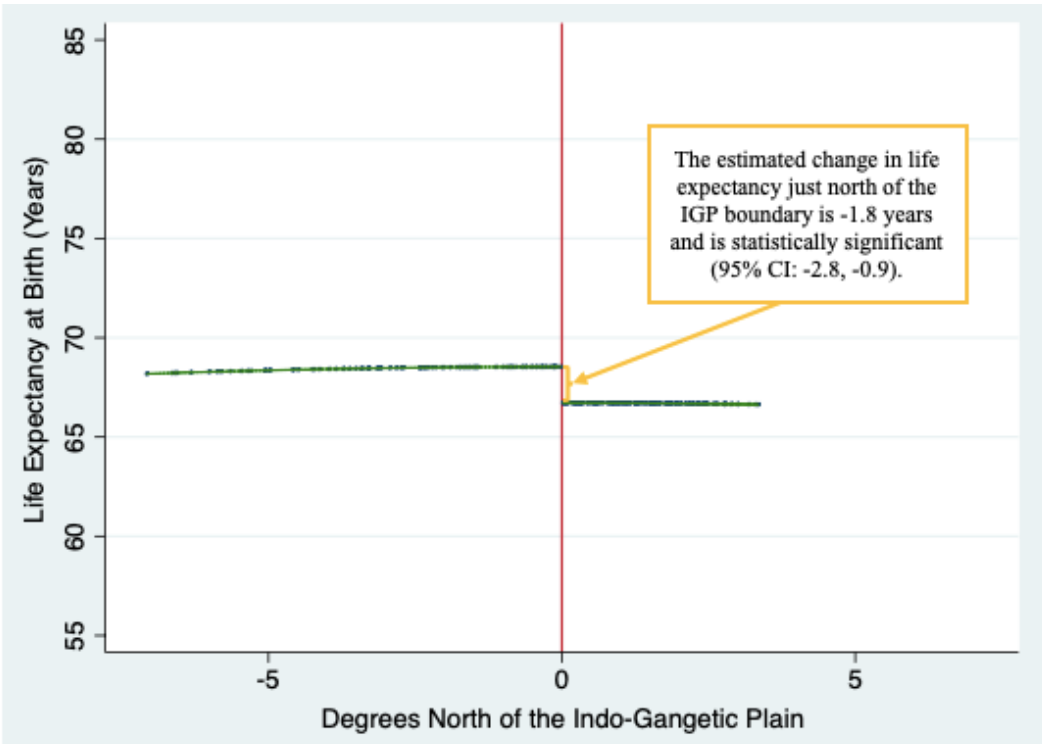

Fig. 3. Fitted values of Life Expectancy at Birth across the IGP boundary

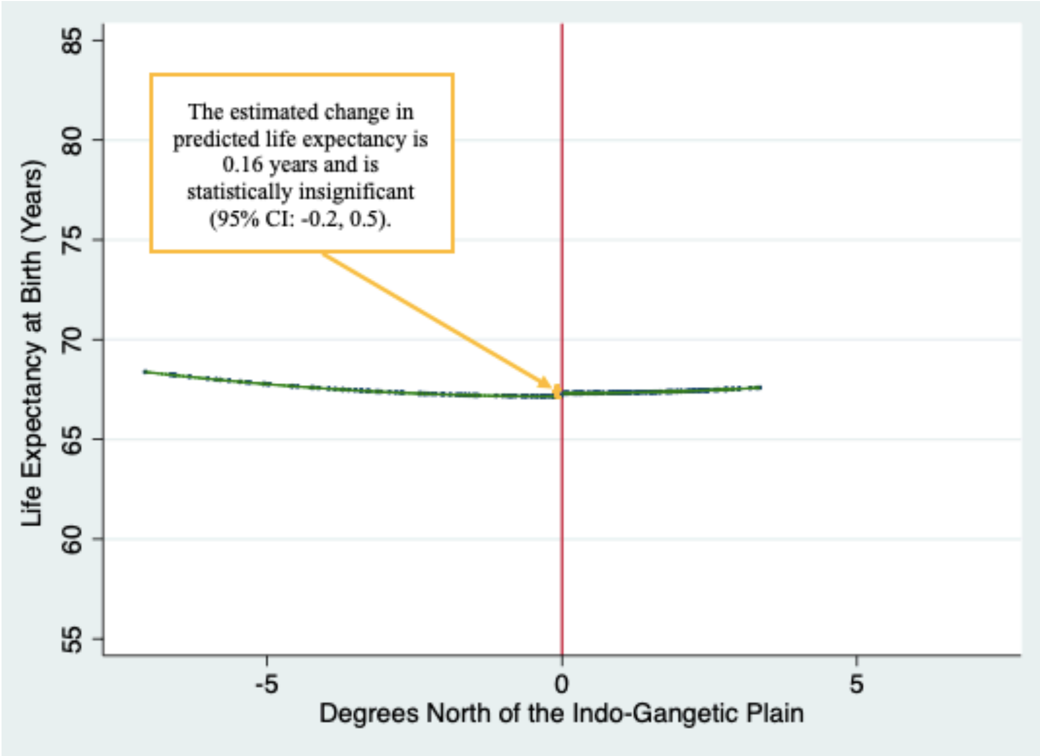

Fig. 4. Fitted values of Predicted Life Expectancy across the IGP Boundary

Regression Results

The presentation of Tables 2 to 4 is explained as follows. Each cell in each of the tables reports results obtained from a separate regression. Each row presents results for a specific health outcome (namely, neonatal mortality rate, infant mortality rate, under-five mortality rate and life expectancy at birth). The columns report results from different model specifications for that particular health outcome. Table 3 has an additional row presenting results pertaining to $PM_{2.5}$.

Table 2 presents the results obtained by using the conventional OLS approach. While Columns (1) and (2) report the regression results by estimating equation (1) excluding and including demographic and socioeconomic variables, respectively. Exposure to $PM_{2.5}$ is found to adversely affect the health status of the population. From column (2), Additional 10 $\mu g/m^3$ of $PM_{2.5}$ exposure raises the number of deaths of new-borns within first 28 days of life by 1.7 per 1000 live births. Similarly, the number of deaths of infants and children under the age of 5 years rises by 3.1 and 3.7 per 1000 live births, respectively, due to additional 10

$\mu g/m^3$ of $PM_{2.5}$ exposure. Life expectancy at birth is found to be negatively affected by exposure to fine particulate matter. A rise in $PM_{2.5}$ exposure by 10 $\mu g/m^3$ reduces life expectancy by 0.54 years. The results are found to be significant at 1%.

Table 2. Impact of $PM_{2.5}$ on health outcomes using conventional strategy (OLS)

Dependent Variable	(1)	(2)
Neo-Natal Mortality Rate	0.09* (0.05)	0.17*** (0.04)
Infant Mortality Rate	0.20*** (0.06)	0.31*** (0.06)
Under-Five Mortality Rate	0.21*** (0.08)	0.37*** (0.07)
Life Expectancy at Birth	-0.04*** (0.01)	-0.05*** (0.01)
Number of observations	234	232
Demographic Controls	No	Yes

Note: Each cell in the table represents a coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses.

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 3 presents the local treatment effect of residing in the IGP on PM2.5 levels and the above four health outcomes using RDD analysis . Each row presents results for a specific outcome, and different columns report results from different model specifications. Column (1) reports the results when the difference in latitude from the IGP boundary is used linearly. Columns (2) and (3) show the impact of “North” when the polynomial in latitude is quadratic. While column (2) excludes demographic and socio-economic controls, column (3) includes them.

$PM_{2.5}$ exposure in districts in the IGP is significantly higher than the districts south of the boundary, ranging between ~15-23 $\mu g/m^3$. Neonatal mortality rate is found to be higher on average by approximately 5.6 neonatal deaths per 1000 live births in the districts lying in the

IGP. The infant mortality rate in the IGP exceeds that in the districts lying below it in the sample by almost 6.5 deaths per 1000 live births. Similarly, the under-five mortality rate is also higher by around 7.4 deaths per 1000 live births in the IGP. This implies that neonatal mortality rate in districts lying above and close to the boundary is around 16.7% higher than the average NNMR in the districts south of the boundary. Similarly, IMR in districts north of the boundary is found to be around 13.6% higher than the districts south of the boundary. A 12.4% higher under-five mortality rate is found in northern districts than the average under-five mortality rate in the relevant southern districts. Life expectancy at birth is found to fall by 2.6 years when individuals reside in the IGP.

Table 3. Impact of being "North" of the boundary on listed variables, RDD

Main Independent Variable →	N_j = 0, 1		
Dependent Variables	(1)	(2)	(3)
$PM_{2.5}$, μg/m^3	15.12*** (2.61)	22.49*** (1.73)	23.33*** (2.01)
Neo-Natal Mortality Rate	3.51 (2.33)	1.80 (1.76)	5.64*** (1.69)
Infant Mortality Rate	1.47 (3.07)	1.18 (2.32)	6.45*** (2.31)
Under-Five Mortality Rate	1.33 (4.18)	-0.22 (3.24)	7.36** (3.14)
Life Expectancy at Birth	-1.26* (0.71)	-1.82*** (0.49)	-2.62*** (0.54)
Number of observations	232	234	232
Demographic Controls	Yes	No	Yes
Polynomial in latitude	Linear	Quadratic	Quadratic

Note: Each cell in the table represents a coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses. Models in column (1) are estimated with a linear control for latitudinal difference and include demographic controls. Models in column (2) include a quadratic in latitudinal difference and do not include demographic and socioeconomic controls. Models in column (3) include demographic and socioeconomic controls along with quadratic latitudinal difference.

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 4 presents results of the 2SLS regression analysis and the organization of the table is similar to Table 3. A rise in $PM_{2.5}$ exposure by 10 μg/m^3 raises NNMR and IMR by 2.6 and 3, respectively. The neonatal period of death constituted around 87% of the infant deaths due to the rise in $PM_{2.5}$. The under-five mortality rate also rose by 3.6 with a 10 μg/m^3 rise in $PM_{2.5}$ exposure. Life expectancy is found to fall by 1.2 years due to a 10 μg/m^3 rise in $PM_{2.5}$, ceteris paribus. We find that the RDD and the 2SLS estimates are much higher than the OLS estimates.

Table 4. Impact of $PM_{2.5}$ exposure on health outcomes, 2SLS

Main Independent Variable →	Fitted values of $PM_{2.5}$ (μg/m^3)		
Dependent Variable	(1)	(2)	(3)
Neo-natal Mortality Rate	0.27* (0.16)	0.08 (0.08)	0.26*** (0.07)
Infant Mortality Rate	0.14 (0.20)	0.05 (0.10)	0.30*** (0.09)
Under-Five Mortality Rate	0.15 (0.27)	-0.01 (0.14)	0.36*** (0.13)
Life Expectancy at Birth	-0.09* (0.05)	-0.08*** (0.02)	-0.12*** (0.02)
Number of observations	232	234	232
Demographic Controls	Yes	No	Yes
Polynomial in latitude	Linear	Quadratic	Quadratic

Note: Each cell in the table represents a coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses. Models in column (1) are estimated with a linear control for latitude. Models in column (2) include a quadratic in latitude. Models in column (3) additionally

include demographic controls. Two observations are excluded in the second and fourth columns because of missing data on income in Palwal (Haryana) and Pratapgarh (Rajasthan).
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Life years saved on attaining air quality standards

In our data, the range of $PM_{2.5}$ exposure is 33-108 μg/m^3 and, the exposure within one standard deviation ranges between 45 and 82 μg/m^3. This is similar to the range of $PM_{2.5}$ exposure in the entire country. Hence, the causal relationship established between air pollution and life expectancy is derived from a sample with distribution of $PM_{2.5}$ exposure that closely resembles that of the country. Hence, it is reasonable to extend these results to rest of India.

When a region is able to attain the national or WHO standards for $PM_{2.5}$, the life years saved can be calculated by using the formula below:

$$Life\ years\ saved\ =\ 0.12\ *\ (Average\ PM_{2.5}\ exposure — \ PM_{2.5}\ standard)$$

where 0.12 is the estimated coefficient of $\widehat{PM_j}$ as reported in Table 4, column (3). In estimating the gain in life expectancy, a linear association between $PM_{2.5}$ exposure and life expectancy is assumed throughout the distribution of $PM_{2.5}$. The assumption bears support from the literature (GBD, 2016; Greenstone and Fan, 2018). In *Appendix* Fig. A1 and Table A1, gains in life expectancy at birth are reported for different states and regions, respectively. The average $PM_{2.5}$ exposure in a region is estimated as its weighted average by using population as the weight. Assuming that the 2014 pollution concentration levels sustain, we find that India can raise life expectancy on average by 1.7 and 5.3 years on meeting the national and WHO standards for $PM_{2.5}$, respectively (*Appendix*, Table A1).

Robustness Checks

The findings are subjected to several robustness checks. First, a variable on interaction between the dummy variable, N_j and the quadratic function, $f(L_j)$ variable is added to the

RDD and the 2SLS (*Appendix*, Table A2). Second, elevation above sea level and altitude of a district are included separately as proxies for topography. The estimates for pollution, early life mortality and life expectancy remain significant at 1% in each case (*Appendix*, Tables A3 and A4). Third, the sample is restricted to different bandwidths of latitudinal difference. The estimates for pollution, early life mortality and life expectancy remain significant even when the bandwidth reduces to 5°, 4° and 3° (*Appendix*, *Tables* A5 and A6).

CONCLUSION

The study attempts to understand the causal impact of sustained exposure to $PM_{2.5}$ exposure on four health outcomes (NNMR, IMR, U5MR and LEB) in India. This is the first long-term exposure study which utilizes district-level data from within the country, unlike the previous studies that have used estimates derived from other countries. The 2SLS estimate (= - 0.12) is more than double the OLS estimate found in the study (= - 0.05). This suggests that the OLS underestimates the causal health impact of air pollution. This may be attributed to the lack of counterfactual leading to omitted variable bias and hence, endogeneity.

The estimated impact of exposure to $PM_{2.5}$ on life expectancy in this study is 71% higher than the estimated value by Pope et al. (2002) for the US (= -0.07). This strengthens the argument that utilizing the estimates from developed countries may not be desirable for developing countries due to several physiological, geographical, socioeconomic, weather-related and other differences between them. Comparing our results with Greenstone and Fan (2018), the estimates from this study are 22% higher as life expectancy reduces by 0.98 years for every additional 10 $\mu g/m^3$ of $PM_{2.5}$ above the $PM_{2.5}$ standard in their study, whereas it reduces by 1.2 years in our study. The IMR effect of change in $PM_{2.5}$ by 10 $\mu g/m^3$ is 6.1% in our study, while it is 8.8% and 9.2% in the studies conducted in Mexico and Africa, respectively (Arceo et al., 2016; Heft-Neal et al., 2018).

Analysing individual cities of India, Delhi, Patna and Allahabad are likely to gain immensely in terms of life years (life expectancy gains exceeding 7 years) if the WHO standard for $PM_{2.5}$ is attained in the districts (*Appendix*, Table A7). The $PM_{2.5}$ exposure has been the highest in Delhi. Delhi has suffered a loss of 1.9 years of life due to increased

$PM_{2.5}$ exposure levels as compared to 1998 levels. Gains in life expectancy in Delhi are around 10.9 years when the WHO standards are reached and around 7.3 years when the national standards are met in the city. A significantly higher improvement in life expectancy can be seen in the states lying in the Indo-Gangetic Plains owing to higher levels of $PM_{2.5}$ in the region (*Appendix*, Fig. A1).

There remain a few limitations in the study. The study utilises data on infant mortality rate to estimate life expectancy at birth at the district level. More accurate estimates could be achieved if district-level data on life expectancy derived from age-specific mortality rates were available. For measuring $PM_{2.5}$ exposure, a combination of satellite-driven data and data from monitoring stations on $PM_{2.5}$ would have been more comprehensive. However, due to the lack of district-level data for the relevant districts from monitoring stations, this study relies on satellite data. The future work can investigate the causal impact of exposure to air pollution (primarily $PM_{2.5}$, and PM_1) on morbidity and the overall mortality rate. An analysis of cause-specific mortality at the district level would be useful. With the presence of more comprehensive datasets, more elaborate studies can be done for India.

ACKNOWLEDGEMENT

The authors gratefully acknowledge Praffulit Bisht for his assistance in constructing the Indo-Gangetic Plains boundary.

REFERENCES

1. A. Donkelaar *et al.*, Data from “Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors.” *Environ Sci Technol* **50 (7)**, 3762-3772 (2016).
2. A. Ebenstein, M. Fan, M. Greenstone, G. He, M. Zhou, New evidence on the impact of sustained exposure to air pollution on life expectancy from China’s Huai River Policy. *Proc. Natl. Acad. Sci. U.S.A.* **114(39)**, 10384-89 (2017).
3. C.A. Pope, T.B. Richard, J.T. Michael, Lung Cancer, Cardiopulmonary Mortality, and Long-Term Exposure to Fine Particulate Air Pollution. *JAMA.* **287(9)**, 1132–41 (2002).

4. D.L. Thistlethwaite, D.T. Campbell, Regression-discontinuity analysis: An alternative to the ex post facto experiment. *J. Educ. Psychol.* ***51*(6)**, 309-317 (1960).
5. E. Lavigne *et al*., Fine Particulate Air Pollution and Adverse Birth Outcomes: Effect Modification by Regional Nonvolatile Oxidative Potential. *Environ. Health Perspect.* CID:077012 (2018).
6. G. Imbens, J. Wooldridge, *Estimation of Average Treatment Effects Under Unconfoundedness. NBER* (2007).
7. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2016 (GBD 2016) Burden by Risk 1990-2016. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2017.
8. Hecht R, Sinha R. GEOMORPHIC ATLAS OF INDO-GANGETIC PLAINS. Available at http://www.iitk.ac.in/gangetic/index.htm. (2003)
9. J. Carre *et al.*, Does air pollution play a role in infertility?: a systematic review. *Environ. Health* **16**, 82 (2017).
10. J. Shah, T. Nagpal, *Urban AQMS in Asia, Greater Mumbai report.* World Bank Technical Paper No 381, World Bank, Washington, DC (1997a).
11. J.C. Aitchison, A.M. Davis, When and where did India and Asia collide?. *J. Geophys. Res.* **112**, B05423 (2007).
12. J.S. Apte, P. Pant, Toward cleaner air for a billion Indians. *Proc. Natl. Acad. Sci. U.S.A.* **116(22)**, 10614-16 (2019).
13. K. Lvovsky, Economic Costs of Air Pollution with special reference to India. Prepared for National Conference on Health and Environment on July 7-9, 1998 at Delhi, India.
14. M. Greenstone *et al.*, Lower Pollution, Longer Lives. *EPW* **L(8)**, 40-46 (2015).
15. M. Greenstone, C.Q. Fan, Introducing the Air Quality Life Index: Twelve Facts about Particulate Air Pollution, Human Health, and Global Policy. Energy and Policy Institute, University of Chicago (2018).
16. M. Kandlikar, G. Ramachandran, The causes and consequences of particulate air pollution in urban India: a synthesis of the science. *Annu. Rev. Energ. Environ.* **25**, 629-684 (2000).
17. M.L. Cropper, N.B. Simon, A. Alberini, P.K. Sharma, *The Health Effects of Air Pollution in Delhi, India*. Policy Research Working Paper No. 1860, World Bank, Washington DC (1997).

18. P. Nema, S.K. Goyal, "Estimation of health impacts due to PM10 in major Indian cities" in *Air Pollution: Health and Environmental Impacts*, Taylor & Francis Group. (2010), pp. 297-310.
19. R. Kesarwani, Estimation of Life Expectancy from Infant Mortality Rate at District Level. *IRJSSH.* **4(6),** 52-63 (2015).
20. S. Burrard, On the Origin of the Indo-Gangetic Trough, Commonly called the Himalayan Foredeep. *Proc. R. S. Lond. Series A, Containing Papers of a Mathematical and Physical Character* **91(628)**, 220-238 (1915).
21. S. Chowdhury, S. Dey, Cause-specific premature death from ambient PM2.5 exposure in India: estimate adjusted for baseline mortality. *Environ. Int.* **91**, 283–290 (2016).
22. S. Heft-Neal, J. Burney, E. Bendavid, M. Burke, Robust relationship between air quality and infant mortality in Africa. *Nature.* **559**, 254-58 (2018).
23. S. Tanaka, Environmental Regulations on air pollution in China and their impact on infant mortality. *J. Health Econ.* **42**, 90-103 (2015).
24. S.K. Guttikunda, B.R. Gurjar, Role of meteorology in seasonality of air pollution in megacity Delhi, India. *Environ. Monit. Assess.* **184(5)**, 3199-3211 (2012).
25. S.K. Guttikunda, P. Jawahar, Application of SIM-air modeling tools to assess air quality in Indian cities. *Atmos. Environ.* **62**, 551–561 (2012).
26. S.K. Guttikunda, R. Goel, Health impacts of particulate pollution in a megacity-Delhi, India. *Environ. Dev.* **6**, 8-20 (2013).
27. World Health Organisation, India takes steps to curb air pollution. *Bull. World Health Organ.* **94**, 487-88 (2016).
28. World Health Organisation, WHO Global Ambient Air Quality Database, Geneva (2018).
29. Y. Chen, A. Ebenstein, M. Greenstone, H. Li, Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy. *Proc. Natl. Acad. Sci. U.S.A.* **110(32)**, 12936-41 (2013).
30. Y. Zhang *et al.*, Chemical composition and sources of PM1 and PM2.5 in Beijing in autumn. *Science of The Total Environment.* **630**, 72-82 (2018).

Appendix For:

Health Effects of Sustained Exposure to Fine Particulate Matter: Evidence from India

Yashaswini Saraswat [1] & Sangeeta Bansal [1]

This PDF file includes :

Definitions
Figure A1
Tables A1 to A7

[1] Centre for International Trade and Development, Jawaharlal Nehru University, Delhi-67

DEFINITIONS

1. Neonatal Mortality Rate (NNMR): Probability of dying during the first 28 days of life, expressed per 1,000 live births.
Source: Computed as per UNICEF definition from NFHS 4, 2015-16

2. Infant Mortality Rate (IMR): Probability of dying between birth and exactly 1 year of age, expressed per 1,000 live births.
Source Computed as per UNICEF definition from NFHS 4, 2015-16

3. Under-Five Mortality Rate (U5MR): Probability of dying between birth and exactly 5 years of age, expressed per 1,000 live births.
Source: Computed as per UNICEF definition from NFHS 4, 2015-16

4. Life Expectancy at Birth (LEB): Average number of years that a new-born is expected to live if current mortality rates continue to apply.
Source: Computed as per WHO definition from NFHS 4, 2015-16

5. Latitudinal difference: Latitudinal difference between a district and a corresponding district lying in the lower boundary of the Indo-Gangetic Plains.

6. Literacy Rate: Proportion of people aged 7 and above who can both read and write with understanding in any language.
Source: Census 2011

7. Access to treated tap water: Proportion of households who have access to treated tap water within their premises.
Source: Census 2011

8. Access to clean cooking fuel: Proportion of households who use LPG or electricity as the primary source of fuel for cooking purposes.
Source: NFHS 4, 2015-16

9. Share of rural households: Proportion of households not in urban areas. An urban area is defined as:
(a) all places with a Municipality, Corporation or Cantonment or Notified Town Area
(b) all other places which satisfied the following criteria:
    (i) a minimum population of 5,000.
    (ii) at least 75% of the male working population was non-agricultural.
    (iii) a density of population of at least 400 sq. Km. (i.e. 1000 per sq. Mile)
Source: Census 2011

10. Consumption expenditure: Household Consumer Expenditure (HCE) is most easily understood as expenditure incurred by households on consumption goods and services, i.e., on goods and services used for the direct satisfaction of individual needs and wants or the collective needs of members of the community and not for further transformation in production.
Source: NSSO, Round 68, 2011-12

11. Share of minority population: Proportion of households belonging to Scheduled Caste, Scheduled Tribe or Other Backward Class.
Source: NFHS 4, 2015-16

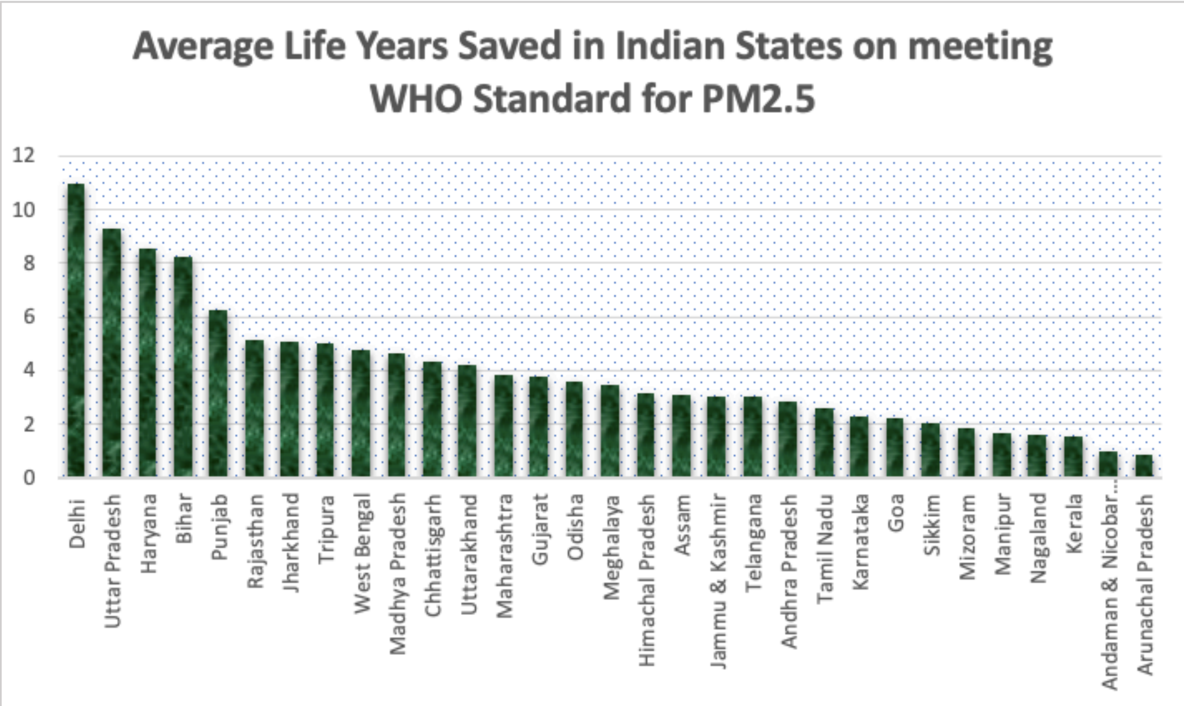

Fig. A1. State-wise gains in life expectancy on meeting WHO standard for $PM_{2.5}$

Table A1. Life years saved on attaining the air quality standards

Region	Weighted Average $PM_{2.5}$ ($\mu g/m^3$)	Life Years Saved: National Standard	Life Years Saved: WHO Standard
National	54.07	1.69	5.29
Indo-Gangetic Plains	83.70	5.24	8.84

Table A2. Inclusion of interaction variable for robustness check

Main Independent Variable →	N_j = 0, 1	Fitted values of $PM_{2.5}$ ($\mu g/m^3$)
Dependent Variable	(1)	(2)
$PM_{2.5}$, $\mu g/m^3$	22.76*** (2.26)	-
Neo-natal Mortality Rate	4.08** (1.84)	0.20** (0.08)
Infant Mortality Rate	3.77 (2.48)	0.19* (0.10)
Under Five Mortality Rate	3.98 (3.35)	0.21 (0.14)
Life Expectancy at Birth	-1.69*** (0.56)	-0.08*** (0.02)
Number of observations	232	232
Demographic Controls	Yes	Yes
Polynomial in latitude	Quadratic	Quadratic
Estimation method	RDD	2SLS

Note: Each cell in the table represents a coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses. The estimates in column (1) denote the impact of being located in the IGP on the relevant dependent variables. The estimates in column (2) denote the change in the dependent variable due to a unit change in the $PM_{2.5}$ concentration level, ceteris paribus.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table A3. Inclusion of elevation for robustness check

Main Independent Variable →	N_j = 0, 1	Fitted values of $PM_{2.5}$ ($\mu g/m^3$)
Dependent Variable	(1)	(2)
$PM_{2.5}$, $\mu g/m^3$	20.90*** (2.40)	-
Neo-natal Mortality Rate	5.99*** (1.96)	0.31*** (0.10)
Infant Mortality Rate	8.78*** (2.74)	0.45*** (0.12)
Under Five Mortality Rate	9.70*** (3.70)	0.51*** (0.17)
Life Expectancy at Birth	-3.04*** (0.64)	-0.15*** (0.03)
Number of observations	228	228
Demographic Controls	Yes	Yes
Polynomial in latitude	Quadratic	Quadratic
Estimation method	RDD	2SLS

Note: Each cell in the table represents a coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses. The estimates in column (1) denote the impact of being located in the IGP on the relevant dependent variables. The estimates in column (2) denote the change in the dependent variable due to a unit change in the $PM_{2.5}$ concentration level, ceteris paribus.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table A4. Inclusion of altitude for robustness check

Main Independent Variable →	N_j = 0, 1	Fitted values of $PM_{2.5}$ (μg/m^3)
Dependent Variable	(1)	(2)
$PM_{2.5}$, μg/m^3	19.46*** (2.56)	-
Neo-natal Mortality Rate	7.05*** (2.21)	0.39*** (0.12)
Infant Mortality Rate	10.02*** (3.12)	0.55*** (0.15)
Under Five Mortality Rate	11.46*** (4.25)	0.64*** (0.20)
Life Expectancy at Birth	-3.04*** (0.70)	-0.16*** (0.04)
Number of observations	232	232
Demographic Controls	Yes	Yes
Polynomial in latitude	Quadratic	Quadratic
Estimation method	RDD	2SLS

Note: Each cell in the table represents a coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses. The estimates in column (1) denote the impact of being located in the IGP on the relevant dependent variables. The estimates in column (2) denote the change in the dependent variable due to a unit change in the $PM_{2.5}$ concentration level, ceteris paribus.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table A5. Impact of "North" on listed variables, RDD for different bandwidths

Main Independent Variable →	$N_j = 0, 1$		
Dependent Variable	(1)	(2)	(3)
$PM_{2.5}$, μg/m^3	22.52*** (2.03)	22.39*** (2.03)	22.38*** (2.11)
Neo-natal Mortality Rate	5.83*** (1.72)	5.92*** (1.73)	6.07*** (1.75)
Infant Mortality Rate	6.85*** (2.36)	6.84*** (2.37)	7.05*** (2.42)
Under Five Mortality Rate	7.58** (3.23)	7.68** (3.25)	8.16** (3.28)
Life Expectancy at Birth	-2.85*** (0.55)	-2.76*** (0.54)	-2.43*** (0.54)
Number of observations	218	211	192
Demographic Controls	Yes	Yes	Yes
Polynomial in latitude	Quadratic	Quadratic	Quadratic
Bandwidth	5°	4°	3°

Note: Each cell in the table represents a coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table A6. Impact of $PM_{2.5}$ on listed variables, 2SLS for different bandwidths

Main Independent Variable →	Fitted values of $PM_{2.5}$ ($\mu g/m^3$)		
Dependent Variable	(1)	(2)	(3)
Neo-natal Mortality Rate	0.36*** (0.11)	0.36*** (0.12)	0.31** (0.12)
Infant Mortality Rate	0.52*** (0.15)	0.52*** (0.15)	0.42*** (0.16)
Under Five Mortality Rate	0.61*** (0.20)	0.61*** (0.21)	0.51** (0.22)
Life Expectancy at Birth	-0.15*** (0.04)	-0.14*** (0.04)	-0.13*** (0.04)
Number of observations	218	211	192
Demographic Controls	Yes	Yes	Yes
Polynomial in latitude	Quadratic	Quadratic	Quadratic
Bandwidth	5°	4°	3°

Note: Each cell in the table represents a coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table A7. 20 Most Populous Districts

District	Population (Millions)	$PM_{2.5}$ Concentration, 2014 ($\mu g/m^3$)	$PM_{2.5}$ Concentration, 1998 ($\mu g/m^3$)	Increase in Life Expectancy if district meets National Standard (40 $\mu g/m^3$)	Increase in Life Expectancy if district meets WHO Standard (10 $\mu g/m^3$)	Change in Life Expectancy Due to Change in $PM_{2.5}$, 1998-2014 (years)
Delhi	17	101.2	85.5	7.3	10.9	-1.9
Thane, Maharashtra	11	41.5	32.9	0.2	3.8	-1
North 24 Parganas, West Bengal	10	43.5	37.1	0.4	4	-0.8
Bangalore Urban, Karnataka	9.6	29.7	27.2	0	2.4	-0.3
Mumbai (Suburban), Maharashtra	9.4	45.2	36.3	0.6	4.2	-1.1
Pune, Maharashtra	9.4	44.8	33.3	0.6	4.2	-1.4
South 24 Parganas, West Bengal	8.2	41.8	37.8	0.2	3.8	-0.5
Barddhaman, West Bengal	7.7	50.7	49.6	1.3	4.9	-0.1
Ahmadabad, Gujarat	7.2	44.2	47.1	0.5	4.1	+0.3
Murshidabad, West Bengal	7.1	57.9	48.9	2.1	5.7	-1.1

Jaipur, Rajasthan	6.6	53.8	50	1.7	5.3	-0.5
Nashik, Maharastra	6.1	36.8	27.9	0	3.2	-1.1
Surat, Gujarat	6.1	39.9	34	0	3.6	-0.7
Allahabad, Uttar Pradesh	6	72.8	65.4	3.9	7.5	-0.9
Paschim Medinipur, West Bengal	5.9	46.2	47.4	0.7	4.3	+0.1
Patna, Bihar	5.8	84.9	63.2	5.4	9	-2.6
Hugli, West Bengal	5.5	47.3	46.6	0.9	4.5	-0.1
Rangareddy, Telangana	5.3	34.6	29.8	0	3	-0.6
East Godavari, Andhra Pradesh	5.2	36.1	30.1	0	3.1	-0.7
Nadia, West Bengal	5.2	53.2	45.4	1.6	5.2	-0.9