Early Life Exposure to Pollution: Effect of Seasonal Open Biomass Burning on Child Health in India

Prachi Singh [1], Sagnik Dey[2] & Sourangsu Chowdhury[3]

[1]Indian Statistical Institute, Delhi
[2]Indian Institute of Technology, Delhi
[3]Indian Institute of Technology, Delhi

Early Life Exposure to Pollution:
Effect of Seasonal Open Biomass Burning on Child Health in India

Abstract

This paper examines effect of outdoor air pollution on child health in India by combining satellite PM2.5 data with geo-coded Demographic and Health Survey of India(2016). Pollution levels vary due to seasonal open biomass burning events (like crop-burning and forest fires) which are a common occurrence. Our identification strategy relies on spatial and temporal differences in these biomass burning events to identify the effect air pollution on child health. Our results indicate that children exposed to higher levels of PM2.5 during their first trimester and during the post-natal period of first three months after birth have lower Height-for-age and Weight-for-age; the effect is not limited to just rural areas, but prominent for Northern states of India which have higher incidence of such events.

JEL Classification: 012 I15 Q53 Q56

Corresponding author:
Prachi Singh
Economics and Planning Unit
Indian Statistical Institute
New Delhi
Email: prachi.singh12r@isid.ac.in

Early Life Exposure to Pollution: Effect of Seasonal Open Biomass Burning on Child Health in India

[PRELIMINARY DRAFT]

1 Introduction

Pollution in any form air or water poses an environmental risk to the health of the exposed population. Literature from both developing and developed nations informs about health effects that air (Chay and Greenstone, 2003) and water pollution (Brainerd and Menon, 2014) have on children. In spite of enormous evidence about ill effects of pollution there is little effective management or regulation present to curtail activities which contribute to high levels of pollution in developing nations.

According to WHO global air pollution database, out of 15 most polluted cities in the world 14 belong to India. Another recently published report by Health Effects Institute on air pollution in India (2018) reports that air pollution was responsible for 1.1 Million deaths in India in 2015. The major contributors to air pollution in India are household burning emissions, coal combustion, agricultural burning and transport. In presence of ineffective pollution regulating policies air pollution levels reach alarming levels in various parts of India. This warranties a closer look at the air pollution problem from the standpoint of welfare of the younger generation currently being exposed to harmful pollutants with possible long lasting effect on their health.

This paper examines the effect of outdoor air pollution on child health in India. Particularly we study the effect of early life exposure to air pollution (as measured by PM 2.5) on child's underweight and stunting measures. We use gridded satellite data on PM2.5 and combine it with the DHS-4, 2016 round for India. Using GPS locations of sampled clusters we are able to produce rich geo-spatial information about local pollution levels in the place of residence of the child. The pollution levels are affected by the open biomass burning events (we call them fire events) like crop burning and forest fires. These open biomass burning events are seasonal in nature as they are driven by cropping patterns and climatic conditions. Most of the crop burning activity is attributable to rice crop residue burning which happens in the months of October and November (Wheat crop residue burning is another major contributor which takes place in the months of April and May). Forest fires also form a big component of open biomass burning in India and it mainly takes place in the dry and hot months of March to May. We are able to link the occurrence of these fire events to local pollution levels using geo-coded NASA's data on global fire-events. In a regression of child health on local pollution level, household income and behavioural choices are omitted variables which make local pollution levels endogenous. We use neighbouring (or non-local) fire-events as an instrument for local pollution levels as they are not related to household behavioural choices or local economic activity, but they affect local pollution levels as smoke and pollutants from these neighbouring fire events can travel long distances. We exploit spatial and temporal variation in these fire-events to identify the effect of outdoor pollution on child health.

Our analysis shows that air pollution negatively affects children's health. Exposure to air pollution during the first trimester decreases both Height-for-age and Weight-for-age for children aged below 5 years. Similar effect on child health is also seen for exposure to pollution during first three months after birth. The effect is present for both urban and rural areas, with Northern states being more vulnerable due to high incidence of fire events and consequent high pollution levels in the areas.

Evaluating this link between poor air quality and child health is important as many regions in India have very high pollution levels which breach the safe standards often. Since stunting is affected by early life exposure to pollution so it can have long lasting effect on life earnings of a child due to poor cognition and it also increases the vulnerability towards hypertension and diabetes. In absence of proper regulation regarding crop burning or effective forest fire management policies in India the problem will only grow in magnitude overtime.

The literature linking air pollution to child health has mostly focused on child mortality.

We instead steer our analysis towards child's growth indicators conditional on child's survival. Most of the studies for developing nations lack rich geo-spatial data for analysis and relies on pollution measures which are estimated for a large area (hence they are riddled with measurement error). To the best of our knowledge this is the first study for India which links child health with local pollution levels and occurrence of fire events. Also open biomass burning problem is well known in India but understudied in terms of its effect on child health, our study is an attempt to bridge this gap.

The paper follows the following structure. The next section provides a literature overview of effect of pollution on child health. Section 3 focuses on the pattern and reasons behind of biomass burning in India. Section 4 describes the various datasets (like NASA's fire events data, satellite pollution data) that we use in our analysis. The next section presents the empirical methodology that we follow and it is followed by results in Section 6. Lastly section 7 concludes with an estimate of the extent of the problem and discusses current state of policies regarding biomass burning in India.

2 Previous Literature

Our work is motivated by the "fetal origins" hypothesis (Douglas and Currie, 2011), which states that the *in-utero* period of a child critically determines mortality outcomes, disease prevalence and future health outcomes, abilities and earnings. The health of a child is a function of both genetic (nature) and non-genetic (nurture) factors. Our genes constantly interact with the environment and exposure to pollution can mutate the genetic coding. These mutations due to interaction with environment are called *epigenetic changes* and the normal functioning of a gene is altered by these changes. During early phase of a pregnancy an embryo contains stem cells which are "coded" to later on develop into different body organs/parts like lung cells, liver cells, teeth etc. Exposure to pollution at this stage can trigger a series of epigenetic changes due to which future health outcomes of a child can be negatively affected.

The intrauterine period has been the focus of many studies in economics literature which have linked occurrence of early life shocks to multiple outcomes. Early life shocks studied in economics literature include incidence of a) disastrous events (like famines, war, drought); b) nutritional shocks (like introduction of iodized salt, pregnancy during Ramadan) and c) pollution (air or water). Currie and Vogl (2013) provide a review of these early life shocks (a and b) on various outcomes; broadly summarized these shocks negatively affect adult cognition, years of schooling, literacy status, adult height and stunting measures; and increase the likelihood of presence of birth defects, prevalence of heart disease and obesity.

The focus of our study is in-utero exposure to air pollution and Currie et. al (2014) reviews landmark studies which have been done in this area. Most of these studies are from developed nations with few exceptions. Similar to previous studies a major part of the literature focuses on learning outcomes (test-scores) and earnings which are negatively affected due to in-utero exposure to pollution (Bharadwaj et al., 2013; Isen et al., 2013 & Sanders, 2012).

The strand of literature which is most relevant for our study has mainly looked at the effect of in-utero or early life exposure to air-pollution on infant mortality and birth weight. Few papers in this area have used natural experiments to causally identify the effect of air pollution on infant survival, for example Chay and Greenstone (2003a and 2003b) use introduction of environmental regulations under Clean Air Act, 1970 and recession in 1981-82 in United states to show that reduction in pollution levels led to reduction in infant mortality. Currie and Walker (2011) show that introduction of congestion reducing automated toll payment systems in United States (which reduced number of idle vehicles emitting harmful pollutants) reduced pre-mature and low birth-weight births. Currie and Neidell (2005) uses spatial and temporal variation in CO levels to analyse the effect of CO levels on infant mortality.

The paper by Greenstone and Hanna (2014) is one of the landmark studies which analyses the effect of water and air pollution regulation policies on infant mortality in a developing nation context (India). Another study from a developing nation includes Foster et al. (2009) which uses Mexico's clean industry certification program to study its effect on pollution (we use a similar measure of pollution i.e. satellite data on Aerosol Optical Depth to infer PM2.5

levels) and resulting respiratory related infant deaths. Wildfires and their negative health effects (like increase in infant mortality, reported asthma cases, pre-term births etc) have also been studied in context of Indonesian wildfire of 1997 (Jayachandran, 2009; Rukumnuaykit, 2003; Kunii et al., 2002; Frankenberg et al., 2005 & Barber and James, 2000), California wildfires (Holstius et al.,2012) and Australian wildfires (O'Donnell and Behie, 2015). Two studies on India inform about the effect of water pollution on child health. Brainerd and Menon (2014) have focused on use of fertilizers in India during crop sowing season which increases concentration of harmful chemicals in water. They find that exposure to these pollutants during the month of conception increases infant mortality and reduces Height-for-age and Weight-for-age for children. Do et al. (2018) have shown that regulation targeting industrial pollution in the Ganga River, led to reduction in water pollution levels and infant death.

In case of India literature has shown how air pollution can affect infant mortality but no study has focused on post-natal growth outcomes for surviving children. Also due to lack of local air pollution monitoring systems which cover the entire nation, there has been no study which calculates local pollution levels and assesses its effect on health outcomes. To the best of our knowledge our study is the first one to use satellite data to calculate local pollution levels (mean pollution levels in 50 km radius) during the in-utero period of a child and study its effect on Height-for-age and Weight-for-age measures in the developing country setting.

3 Background

India has a substantial amount of land under cultivation(60%) and under forest cover(25%), with majority biomass burning events taking place in these areas. Over the past few decades Indian agriculture has been marked with expansion of irrigation facilities, adoption of high yield variety seeds and increased mechanisation (like use of combine harvester). A combination of these factors led to adoption of multi-cropping system by farmers which leaves little time in between the harvest of one crop and sowing of another. In this scenario crop residue burning thus emerged as the quickest and cheapest way to get the farm ready for the next crop. Cereals are the prime contributor to crop burning activity in India, with rice and wheat crop residue burning forming the major chunk of residue burning process (Jain et al, 2014). Two major residue burning seasons are thus related to crop harvest seasons: kharif crop harvest (rice stubble burning) which takes place in the months of October and November; and rabi crop harvest (wheat straw burning) which happens in the months of March to May.

Biomass burning in India is not limited to just crop residue burning, it covers forest fires as well. Forest fires or wildfires are caused by various factors acting in conjunction with each other. These factors together form the basis of a concept called *"Fire Triangle"* according to which mainly three factors are needed to start a wildfire: fuel (biomass in form of vegetation growth), oxygen and appropriate climatic conditions (high temperature, low pressure, windy conditions). In most cases wildfires start in scrub area as a surface fire with low intensity, but with enough fuel build up their intensity increases and they spread quickly to other areas. Rising ground fires posit a threat to tall trees which upon catching fire burn with higher intensity. This is called a *ladder effect* in which ground fires climb surface fuels to spread to canopy and result in an even bigger fire called *Crown fires*. Forest Survey of India lists vulnerable months for each state when forest fires are most likely to happen, which mainly span the high temperature months from March to June. Wildfires happen due to both intentional and unintentional human activity. In North Eastern states and in states along the Eastern ghats slash and burn activity is rampant wherein vegetation in forests is cut (slashed) and then burned to clear the piece of land for human use. In alot of cases unintentional human activities like leaving active cigarette butts behind in open forests lead to forest fires. Other natural factors which cause forest fires include lightening which produces a spark to start a fire in dry vegetation.

In western countries forest fires are mainly responsible for the carbon content release due to biomass burning however in case of India (and other South Asian countries) crop residue burning contributes the most to total carbon release. Figure 1 shows that in South Asia, India stands out both in terms of total area burned (4.5 Million hectares burned in 2015) and in terms of total carbon content (1.5 million metric tonnes) released due to biomass burning (we also include China for comparison, which does worse than India in terms of open biomass burning). Figure 2 provides details about biomass burning events in India (data for years 2010-2016). A raw count of biomass burning events in India shows that roughly both crop residue burning and forest fires contribute equally (50% each). However if we weigh these events based on the population density [4] of the area in which these events occur then we see that crop burning events contribute more to the total biomass burning events (65%). This mainly happens because residue burning activities happen in more populated areas as against forest fires which happen in low density areas. Figure 3 provides the population weighted split between forest fires and crop residue burning in few selected states in India. As can be seen in this graph, with an exception of Punjab almost all other states are affected by both forest fires and residue burning.

Biomass burning is a major source of pollution as it releases harmful pollutants like Carbon Dioxide(CO_2), Carbon Monoxide (CO), Sulphur Oxides and particulate matter (PM) in the atmosphere. The release of harmful pollutants in the atmosphere is captured by aerosol loading [5] in the region. Studies have found that aerosol loading increases in the downwind regions and in the vertical direction as well. Kaskaoutis et. al (2014) find that crop burning in Punjab has an effect on aerosol properties of the Indo-Gangetic Plains, also particulate matter (PM2.5) concentrations increase near ground surface and the concentration of pollutants fall as we move from west to east India. To summarize, fine particulate matter released during biomass burning incidents have long range travel properties and affect not just the local areas but far away regions as well.

In Figure 4 we plot mean PM2.5 and mean number of biomass burning events (mean for all India across all sampled clusters) for each month-year from 2010 to 2016. The fire events graph (dotted line) shows two clear peaks in the summer months (March to May) and winter months (October and November) due to forest fires and crop burning activities which

[4] Geo-coded fire events have been projected onto land mask cover for India to categorize each fire event as an event which happens in a forest area vs cropped area. This data is then projected onto density map of India, to get the density of the population in which these events take place.

[5] Aerosol loading is the suspensions of solids and/or liquid particles in the air that we breathe. Dust, smoke, haze are also part of aerosol loading

happen during these times. The solid line corresponding to mean PM2.5 values follow a similar pattern with higher pollution levels for months with higher fire events incidence and lower pollution levels during the rainy period in monsoon season [6]. We use this relationship between PM2.5 and fire-events to identify the effect of pollution on child health.

4 Data

4.1 Demographic Data

The demographic data we use comes from the fourth round of Demographic and Health Survey (2015-16) for India. The DHS is representative at the national and regional level, it contains detailed information about birth history of each woman who was interviewed. In this latest DHS round for India 601,509 households were interviewed, which included 0.7 million eligible women in the age group 15-49 and 0.22 million children aged below 5 years for whom anthropometric measures of health were collected. The DHS-IV sample is a stratified two-stage sample and the primary sampling units (PSUs) correspond to villages in rural areas and blocks in urban areas. The overall sample comprises of around 28000 clusters and the GIS data for all these clusters is also available.

We use anthropometric measures like Height-for-age and Weight-for-age (WHO standard z-scores) to measure the health effect that outdoor pollution has on child health. We also make use of child characteristics like birth order, age and gender. As informed by previous literature mother's characteristics can also affect child's health and hence we control for the age at which she had the child and her education level. DHS data also contains information about source of water, method of cooking and toilet facility of the dwelling, all of these characteristics have important bearing on health of a child and we include them in our analysis as well. An important feature of the DHS data is the birth history of a woman, the month and year of every child ever born to a woman is recorded. We use the location

[6] Climatic conditions determine pollution levels to a great extent, for example pollution levels tend to be higher during winters as the air is heavier due to low temperature with little or no winds because of which pollutants tend to settle near the ground surface

of the cluster and birth date of a child to construct measures of exposures to fire events at different points of time for the child (first three months after birth and in-utero exposure in the three trimesters separately). We make an important assumption that the place of birth (and place of stay of mother when the child was in-utero) is same as the same of current residence of a child [7].

We provide summary statistics of our analysis sample in Table 1. 52% of children in our full sample are males with mean age around 30 months (2.5 years old) and the mean birth order of children is 2.2. The average age of at which mothers have children is 24.5 years. Mother's had on an average 6 years of education and 97% of fathers are literate. Three-fourth of our sample consists of rural households and the same proportion of households report their religion to be Hindu and the mean household size is 6.5. 88% of the households are headed by a male member and the average age of household head is 44.5 years. 85% of the households have an electrical connection, only 23% of the households use piped water as their source of drinking water, 28% of our sample uses clean source of cooking fuel like LPG or bio-gas and the mean open defecation rate in a cluster is 43%.

The mean Height-for-age and Weight-for-age Z scores for our sample is -1.48 and -1.53 respectively (mean weight-for-height is -0.96 for our sample). Height-for-age is a measure of stunting and it represents the effect of early life shocks that a child receives. Stunting generally occurs before age two and its effects are largely irreversible. It is associated with an underdeveloped brain, with long-lasting harmful consequences, including diminished mental ability and learning capacity, poor school performance in childhood, reduced earnings and increased risks of nutrition-related chronic diseases such as diabetes, hypertension, and obesity in future. Weight-for-age (underweight measure) reflects body mass relative to chronological age. It is influenced by both the height of the child (height-for-age) and his or her weight (weight-for-height). Deaton and Dreze (2009) emphasize on the use of Weight-for-age as the health status indicator for children as its a comprehensive measure which captures both

[7] This assumption is a standard assumption which is employed by many papers which used DHS data for analysis. In our sample the mean number of years for which the interviewed family has stayed at the place of residence is around 15 years. See Brainerd & Menon, 2014

stunting and wasting. The mean level of PM2.5 is above $40\mathrm{ug/m}^3$ during all critical windows of development.

4.2 Pollution Data

In India, ground-based PM2.5 measurement started post 2009 under the national ambient air quality monitoring program maintained by Central Pollution Control Board. The network has slowly expanded to around 90 sites across 35 cities over the years, which leaves majority of India unmonitored. Amongst these cities, only Delhi has greater than 20 monitoring sites while most of the cities have a single monitoring site. Furthermore, most of the sites do not have continuous temporal data. To address the paucity in ground-based PM2.5 data in India, we estimate PM2.5 exposure using satellite data (van Donkelaar et al., 2010). We convert aerosol data retrieved from MISR (Multiangle Imaging SpectroRadiometer) to PM2.5 using a conversion factor. The inferred PM2.5 are regridded at 0.1 * 0.1 degree resolution (10km*10km grid) using spline interpolation technique. The low bias in satellite-derived PM2.5 data is corrected using regression method described in our previous work (Dey et al., 2012). Previously we used this database to estimate premature mortality burden at district level. Here PM2.5 statistics are generated at monthly scale for its use in the regression model. We use cluster location from DHS data and calculate mean PM2.5 in the 50km radius for three month periods at 4 different points of time over life cycle of a child (as fetus: 3 trimesters and post-natal: first 3 months after birth). The data for pollution levels in first trimester are missing for 3.5% of our observations, these missing observations are due to missing satellite retrievals due to cloud covers.

4.3 Fire incidents Data

Our source of biomass burning events (called fire incidents) is NASA's Fire Information for Resource Management System (FIRMS) data which captures real-time active fire locations across the globe. The FIRMS data that we use is called MODIS (shortform for MODerate Resolution Imaging Spectro radiometer) data and it records fire incidents at pixel level where each pixel is identified by a latitude and longitude reading. Each latitude (and longitude) is the center of a 1 km fire pixel (1 km X 1 km in size). This data records not just the location of a fire but also the brightness temperature of fire (in Kelvin units) and date and time when the incident was picked by the Terra satellite. An observation for a fire incident in MODIS data for a latitude and longitude does not necessarily mean that the size of the fire is 1 square kilometre, but it means that atleast one fire is located within this fire pixel (under good conditions the satellite can detect fires as small as $100\mathtt{m}^2$). The MODIS data is available on a daily basis since November 2000 and NASA reports that the fires captured by this dataset are mostly vegetation fires. NASA data on fire incidents also provides a variable "confidence", which depicts the quality of the observations and it ranges from 0-100. Following ecological literature which has explored this dataset, observations with medium confidence level (value greater 70) have been selected for our analysis[8] . We provide our analysis with an even higher cut-off value of 85 and our results essentially remain stable [9]. NASA's FIRMs data can also have some missing values attributable to satellite sensor outage however major incidents reported for sensor outage happened in years 2001-2003 which precedes our analysis period.

We use the cluster GIS information from DHS data and calculate the total number of all fire events which took place in a) 50km radius (we call this local exposure) and b) between 75 and 50 km radii (non-local exposure) for 3 month periods at 4 different points of time over life cycle of a child. To ensure respondent confidentiality, all clusters in the DHS data are displaced from their true location. The displacement is done by displacing an urban cluster by 2km and a rural cluster by 5km with 1% of the rural clusters being displaced by 10km. The displacement can take place in any direction but the cluster remains within the country boundary, within the same state and district. We take the radius for our analysis to be 50 km which is large enough so that the true location of the cluster is contained within the 50km radius circle, additionally a large radius ensures that both near and far fire-incidents get captured (as smoke from fire-incidents which do not happen in the immediate

[8]See Srivastava and Garg, 2013

[9]We also conduct alternative analysis with confidence value as 0 and 85, all our results still hold

neighbourhood can also travel long distances).

The exposure to fire-events and mean PM2.5 variables essentially contain the spatial and temporal variation in our data. They are constructed for each child using information about conception location (cluster $\mathtt{c}$) and date (quarter $\mathtt{q}$ and year $\mathtt{t}$, back calculated using date of birth). Fire events in the neighbourhood affect the local PM2.5 levels. We show this link in Figure 5 an Figure 6. Figure 5 shows a descriptive relationship between PM2.5 levels and fire incidents. We calculate mean number of fire incidents and PM2.5 for each month in 50 km radius for each state and then produce a linear fit of the relationship between the two. As is evident there is a positive relationship between the variables and for comparison we plot few observations as well. Punjab which is depicted by a triangle has the highest number of fire incidents in October (the time of rice crop harvest and wheat sowing season) and correspondingly very high pollution level as well. Manipur (represented by square figure) which is one of North Eastern states has lesser number of fire incidents than Punjab in its peak fire season and lower level of pollution. Other states (Kerala-circle, Madhya Pradesh-diamond) have relatively very low number of fire incidents and much lower PM2.5 levels. In Figure 6 we plot the relationship between local mean PM2.5 and non-local total number of fire-events (which lie between 75 and 50 km radii around cluster location). This shows that open biomass burning events happening in the surrounding areas is positively correlated with local pollution levels. We now link pollution levels during the in-utero period with anthropometric measures for children in figure 7. The descriptive graph shows a negative relationship between Height-for-age and exposure to pollution, we explore this relationship in greater detail later in section 5.

We now focus on the temporal and spatial variation in fire-events. Figure 8 shows the mean number of fire-events across all sampled clusters in India (average for data from year 2010 to 2016). The graph shows the seasonal pattern in our data, in the dry and hot months from March to May the number of fire-events increase due to more incidents of forest fires but also because this period coincides with wheat crop residue burning. The months of June to September is the monsoon period when almost no fire-events are captured. October and November are the rice crop harvest months. The rice crop residue burning is primarily responsible for biomass burning during these months. Hence we see a seasonal pattern which is dictated by both weather and crop cycles. Figure 9 shows a similar pattern for select states, the North Eastern states of Assam and Manipur have highest number of fire incidents in the first quarter, Haryana and Punjab show an uptick in fire incidents in quarter 2 and 4 which correspond to the harvest season of wheat crop and rice crop. The southern states have very low number of fire incidents across all months.

Figure 10 shows the spatial variation in our data at the district level (although it should be noted our empirical strategy uses the spatial variation at a much smaller level of a PSU). The figures shows total fire incidents that happen in a district in a year (average for years 2010 to 2016). Districts in Punjab and some in North Eastern states (Assam, Meghlaya, Tripura, Manipur and Mizoram) are high intensity areas. The Western and few parts of Southern of India are almost unaffected by fire incidents while few districts in Central India (Madhya Pradesh and Chattisgarh) and states along the eastern ghats(like Odisha and Andhra Pradesh) show medium intensity of fire-incidents. [10]

5 Methods

We begin by investigating whether early life exposure to outdoor pollution has an impact on child health. We focus on in-utero exposure to outdoor pollution during the first trimester. Formally, we estimate a fixed effects regression as described below :

$$H_{icqt} = \theta_1 PM_{cqt} + \beta X_{icqt} + \gamma_c + \delta_t + \lambda_q + \rho^1_{ct} + \rho^2_{qt} + \varepsilon_{icqt} \tag{1}$$

where PM_{cqt} captures the mean PM2.5 in the 50km radius for the first trimester for a child who is conceived in cluster c. Quarter q and year t correspond to quarter and year of conception. Different clusters (villages or blocks) can have different levels of development (health infrastructure) which can affect health of a child hence we include cluster fixed effects

[10] As discussed in section 3 each state has a different mix of open biomass burning, which comprises of crop-burning and forest-fires. Punjab has all reported fires as crop fires, and other states like Assam, Madhya Pradesh, Chattisgarh etc have around 50-65% of total fires as crop fires and the rest are forest fires.

in our specification. We also include quarter fixed effects and year fixed effects to account for any heterogeneity present at this level. We next introduce two more interactive fixed effects: cluster into year, ρ^1_{ct} and quarter into year, ρ^2_{qt}. These fixed effects capture the spatial and temporal variation in our data. Our identification comes from these fixed effects, that is essentially we compare children who are born in the same cluster and year but in different quarters. The children being compared thus have varying exposure to outdoor pollution as some quarters have a high levels pollution while others have low pollution. Simultaneously we also compare children across different clusters but who are born in the same quarter and year, this exploits the spatial variation in pollution levels across clusters.

Our main outcomes of interest (H_{icqt}) are z-scores for Height-for-age (stunting measure) and Weight-for-age (underweight measure) for children below 5 years of age. We also control for other confounding factors in the vector X_{icqt} which includes gender, birth order and age of child, mother's and father's educational status, mother's age at birth, age and gender of household head, dummy for whether household has pipedwater, has clean cooking source, whether household practices open defecation and the fraction of households who practice open defecation in the cluster (excluding self).

It is essential to define a *local* area for a household which corresponds to the region of economic activity that a household depends on and also affects based on its behavioural decisions. The economic activity of a household determines key inputs (like income) which feed into the production function of health of a child. An example of this can be dependence of a household on nearby forest resources for fuel-wood consumption or for livelihood (if it sells these resources in a market). In this case the choice of use of fuel-wood by household affects the pollution level in the region. Additionally the forest cover is affected by the demand for forest resources (like fuel-wood) in the market, which in turn affects the pollution level in the area where they are finally consumed. A similar logic holds true for crop residue burning as well, its a conscious decision taken by a household which impacts local pollution levels and at the same time affects farmers income which is a determinant of child health. This essentially points towards the fact that local pollution level is endogenous in the region of economic activity of the household.

In India a little over 40% of households use wood as their source of fuel for cooking (NFHS-4 Report for India) [11]. An active fuel-wood market exists in India and Foster and Rosenzweig (2003) discuss that forest cover will depend on geographic scope of the market for forest products. This geographic scope of the fuel-wood market can inform us about the area of economic activity of households which sell fuel-wood. Chakravorty et. al (2014) show that fuel-wood collection is likely driven not only by rural household demand but especially by demand from towns in close proximity (geographic scope of fuel-wood market). As the proximity to nearest town increases the number of sellers of fuel-wood decreases sharply, the average distance to town for these sellers in their sample from IHDS is around 15 km. We choose a large enough radius of 50km (we also show results for a smaller 30km radius) around the household location to ensure that this local fuel-wood market is contained in it.

The household behavioural choice of collecting fuel-wood or crop-burning and household income are omitted variables in our specification hence the local pollution variable is endogenous. To solve this endogeneity problem we rely on an instrumental variable which in our case is the number of fire-events which happen in non-local areas. In precise terms these are the fire-events which happen between 75 and 50km, they impact local mean PM2.5 levels but they are not affected by household behavioural choices. The IV that we use has been explained diagrammatically in Figure 11, where the light gray center denotes the cluster location, the white circle forms the 50 km radius around the cluster and the grey ring are the fire-events which take place between 75 and 50 km radii around the cluster. Our endogenous variable is the mean PM2.5 variable which is calculated for the white circle and the number of fire-events in the grey ring form the IV. We choose the fire-events in non-local far away areas (75 to 50km radius) to ensure that they belong to a region which is not a part of economic activity area of a household. This essentially removes the effect of dependence on crop-burning or nearby forest resources (or farmlands) for livelihood or fuel-consumption. By capturing fire-events in this ring, we ensure that we only capture the part which contributes to the local pollution levels but is not correlated with household behavioural choices.

[11] This fuel-wood is obtained from forests and the demand for fuel-wood depends on household's income (Chaudhuri et. al, 2002).

Formally our first stage is the following:

$$\mathsf{PM}_{cqt} = \theta_2 \mathsf{Fire}_{cqt} + \eta_c + \sigma_t + \omega_q + \Omega^1_{ct} + \Omega^2_{qt} + \varepsilon_{icqt} \quad (2)$$

In equation (2) all variables are same as the ones which are used in equation(1) [12]. Fire_{cqt} is our IV which is the number of fire-events which happen between 75 and 50km [13]. In this 2SLS methology, the fitted values of local PM2.5 are obtained from equation (2) and then plugged in equation (1) to get unbiased effect of local outdoor pollution on child health. Using this IV we are able to purge the effect of other polluting activity (whose spatial and temporal patterns are not same as the pattern for biomass burning events) and focus on pollution due to non-local fire-events which is independent of household behavioural choices.

There are two main limitations of our study, first we are only able to capture big fire-events as satellite recordings for fire-events fail to capture smaller fires which happen in smaller farms. The satellite can only capture fires which burn brightly enough (big enough to be quite intensive which guarantees that they get captured) to be recorded below the tree canopy or cloud cover. The estimates we thus get can be considered as the lower-bound effect of outdoor pollution due to non-local fire-events on child health. Secondly the perfect IV is the one which captures all non-local fire-events in the *downwind* direction with high enough wind speed. However since we lack a rich enough ecological model which tags each fire-event as a downwind fire-event with high enough wind speed that it affects local pollution levels, so our model is the second best model by which we capture the effect of non-local fire-events on local pollution levels.

A recent paper by Agarwal et al.(2017) cautioned against using month of birth data from DHS due to uneven measurement timing across the interview year which induces differential mean age at measurement across birth month. This differential age at measurement gets translated into a difference in height-for-age (or any other anthropometric measure which is age dependent). To carefully account for this difference in age at measurement we introduce multiple controls for age of a child like linear, quadratic or cubic control and dummies for

[12] Also included are the child, mother , household level controls which are a part of X_{icqt} in equation (1)
[13] In our robustness checks we show that our results are robust to alternate radii specifications.

child age. Agarwal et at. (2017) also show that there might be some measurement error in the month of birth, we address this problem by not focusing on a single month but rather we look at quarter of conception and additionally we don't have any reason to believe that this measurement error will be different across high and low intensity regions (in terms of pollution levels).

6 Results

6.1 Pollution and Child Health

Table 2 presents the estimates for the first stage of 2SLS methodology outlined in the previous section. Columns 1 and 2 presents the results for the first stage of our 2SLS methodology using our main IV variable, that is number of fire-events which lie between 75 to 50 km radius. We introduce quarter, year and cluster fixed effects and their interactions to account for any omitted variables at these levels. As expected the relationship between the endogenous variable - local PM2.5 in 50km radius and our IV is positive and highly significant. This is in line with our hypothesis that particulate matter from fire-events far away affect local pollution levels. The first stage F-stat is also greater than 10 (rule of thumb). These results represent that local PM2.5 variation is affected by the seasonality present in biomass burning events happening in non-local adjacent areas. In columns 3 and 4 we show the effect of local fire-events which happen within 50km radius on local pollution level. We don't believe that this second IV meets the exclusion restriction as fire-events happening locally represent household behavioural choices, also these fire-incidents can destroy local resources which households rely on for their livelihood which in turn can impact their income and hence can have an effect on the inputs which are important determinants for child health.

Table-3 presents the OLS and IV results of effect of outdoor pollution in the first trimester on outcomes. As explained in the previous section the OLS regression of outdoor pollution on child health (equation (1)) is riddled with endogeneity problem, hence the estimates that we see in columns 1 and 2 are biased. We next move to the second stage results obtained using

2SLS strategy. We first note that our rk-LM statistic is highly significant. We find that both weight-for-age and height-for-age are negatively affected by outdoor pollution experienced in-utero during the first trimester. The mean magnitude of the effect of a 10% increase in outdoor pollution in first trimester on HFA-Z(WFA-Z) is -0.013(-0.010) standard deviation units [14]. This is similar to the results which were found by Brainerd and Menon(2014) owing to exposure to water pollutants in the month of conception.

We next provide results for child health outcomes with alternative controls for age of a child since HFA-Z (and WFA-Z) is a function of child's age. Children who are born in the same year are measured at different ages in the DHS sample hence this exercise of modelling WFA-Z (and HFA-Z) with alternate age controls becomes important (Agarwal et.al, 2017). In Table 4 from column 1 to 4 we present WFA-Z results (column 5 to 8 for HFA-Z) with linear, quadratic, cubic and dummy variable controls for child's age. The coefficient of the pollution variable remains significant and quite stable across all the four alternative specifications.

Literature informs that variation in concentration of pollutants is highly correlated as they often emanate from same sources. So although the focus of this study has been on outdoor pollution as captured by PM2.5 but the results that we see might represent the effect of other pollutants as well (like CO, CO_2 and Sulphur Oxides etc).

6.2 Robustness Checks

In this section we provide multiple robustness checks for our results. Local weather condition like rainfall can play an important role as rainfall makes the ash and other pollutant particles settle on the ground thereby reducing pollution levels. In Table 5 (columns 1 and 2) we provide our original result (effect of exposure to outdoor pollution in first trimester on HFA-Z and WFA-Z) but with extended control for local rainfall level in the 50km radius around the cluster. The number of observations is slightly smaller than before due to few random missing rainfall information. Our original results still hold and the magnitude of the effect

[14]Mean effect on HFA of a 10% increase in 1st Trimester = θ_1*0.1* (Mean PM2.5) = -0.307*0.1*0.4248 = -0.013

is larger after controlling for rainfall.

In our identification strategy in equation(1) we define quarter as the quarter which corresponds to the month in which the child was conceived and the exposure to outdoor pollution of a foetus *started*, in our first robustness check we change the way we assign quarter to each child. We alternatively define quarter as the quarter in which the three month exposure to outdoor pollution *ended*. For example we previously we defined quarter as follows: if a child was conceived in month of Feb then the quarter was first, we now alternatively assign quarter as second corresponding to month April that is when the three month exposure period ended for the child. Table 5 (columns 3 and 4) presents the IV results corresponding to this exercise and we see that the effect of outdoor pollution on HFA-Z and WFA-Z is still negative and significant.

The analysis uptil now used fire-events happening in 75 to 50km radius as the IV for local mean PM2.5 in the 50km radius around the cluster location. We now provide results for alternate radii specifications to test the sensitivity of our model. In Table 6, columns 1 and 2 the IV being used is the total number of fire-events in 100 to 50 km radius (the gray ring area has been expanded in Figure 11). In columns 3 and 4 the IV being is the total number of fire-events in 50 to 30 km radius for local mean PM2.5 in the 30km radius (compressing the white inner circle in Figure 11). We find that our results are of similar magnitude and still remain significant. Lastly in columns 5 and 6, we drop the observations corresponding to state Punjab which records highest number of fire-events[15]. This has been done to ensure that our results are not driven in any way by the state of Punjab which is affected by high levels of pollution corresponding to highest level of recorded fire-events in India. Our results become larger in magnitude and are more significant after dropping the state of Punjab.

We check whether our instrument meets exclusion restriction in Table 7. We regress various characteristics of a household (and its members) on our IV, essentially an insignificant result shows that there is no systematic relationship between our IV and household (and its member's) characteristics. The level of education of mother and literacy status of father is not systematically related to our IV. The asset ownership of a household and rich-poor status

[15] Almost 25% of total fire-events in India take place in Punjab.

of a household is also unrelated to our IV. The household size is also not significantly related to fire-intensity in non-local areas. The only exception is the result for polio vaccination, where the children living near high-intensity fire areas are more likely to be vaccinated. Preterm births are also associated with exposure to pollution, however in our sample we find no such relationship.

Do mothers plan conception?

An important threat in our analysis can be avoidance behaviour by mothers, that is if mothers purposely avoid particular months for conception due to their concern about future child health related to seasonal biomass burning activities. We test this by looking at birth history of mothers for last five years, we calculate the number of conceptions in each month for each mother living in cluster c. For example if a mother conceived two kids in two different years in the month of February then the count will be 2 for this variable corresponding to the month of February and it will zero for all other months. We regress this variable (number of conceptions per month) on mean of local and non-local fire-events to see if there is any systematic relationship between the two. We also control for mother's education, father's literacy level, characteristics of household head and wealth index of the household. Month and cluster fixed effects are also introduced to account for any time related and region related heterogeneity. We present these results in Table 8. We find that neither local (fire events in 50 km radius) nor non-local (fire events between 75 and 50 km radii) has any impact on mother's conception behaviour.

6.3 Heterogeneity

We provide disaggregated regressions for height-for-age for the following sub-samples: Rural-Urban and North-South. By splitting our sample into rural and urban sample (see Table 8) we find that the effect is present in both rural and urban areas, with marginal effect being larger for children living in urban areas. This can possibly be due to the fact children in rural households are already disadvantaged and marginal effect of exposure to outdoor pollution is not as large. In comparison the urban kids who live in households with better base level health standard (cleaner cooking fuels as well) have larger marginal effects corresponding to exposure to outdoor pollution. We also provide results by splitting our sample into Northern-Southern states and find that most of the effects that we see are limited to North India.

We now focus on other time windows of critical development, that is second, third trimester and the post-natal period of first three months after birth. Table 10 summarizes our results, we find that in-utero exposure to outdoor pollution which is experienced by the mother (and her foetus) in her second and third trimester has no impact on Height-for-age, but some negative effect is present on Weight-for-age corresponding to exposure in second trimester. Exposure to outdoor pollution during post-natal period also affects child growth negatively. The number of observations is lower as only children above 3 months of age have been retained for analysis. Both height-for-age and weight-for-age are adversely affected by outdoor pollution experienced during first three months after birth, a 10% increase in outdoor pollution decreased HFA-Z and WFA-Z by 0.011 and 0.015 standard deviation units.

7 Conclusion

Outdoor pollution in India breaches safe standards in many areas. We link outdoor pollution to biomass burning which is a significant source of carbonaceous aerosols, it plays a vital role in atmospheric chemistry, air quality, ecosystems, and human health. Our analysis shows that outdoor pollution is affected by neighbouring biomass burning events; which are used to causally infer the effect of outdoor pollution (as measured by PM2.5) on child growth indicators. We find that a ten percent increase in PM2.5 levels during first trimester lead to a reduction in Height-for-age (HFA-Z, stunting measure) and Weight-for-age (WFA-Z, underweight measure) by 0.015 and 0.010 standard deviation units respectively. Our weight-for-age estimate are quite similar in magnitude to the effect which was found in Brainerd and Menon (2014) of exposure to water pollutants during the month of conception. We also find that post-natal exposure during first three months after birth to outdoor pollution reduces HFA-Z (WHA-Z) by 0.012 (0.016) standard deviation units. Figure 12 summarizes our results graphically, exposure to outdoor pollution during different critical windows of growth of a child is associated with worse child health outcomes. All the estimates are negative with significant effect present for exposure to pollution during first trimester, second trimester (only WFA-Z measure) and post-natal period.

The above results establish that exposure to pollution is linked to stunting in childhood. We now provide an estimate of this problem on GDP of India using a back of an envelope calculation based on Galasso et al. (2016) study. This study does a literature review of effect of stunting on GDP. Stunting affects GDP of a nation via three channels: lower returns to lower education, lower returns to lower height and lower returns to lower cognition. For India where 66% of the workforce was stunted in childhood, this study estimates that a complete elimination of stunting would have increased GDP by 10% [16]. We use a point estimate of probability of being stunted due to outdoor pollution, and find that a 10% increase in outdoor pollution leads to a 0.036% reduction in GDP.

India needs effective policies regarding regulation and management of outdoor pollution, since the current policies are currently ineffective. For example the budget allocation for effective management of forest fires is really small and remains unused in every financial year. Similarly the government has committed itself to subsidizing the use of happy-seeder technology (this is an alternative to combine harvester, it leaves rice residue in form of a mulch on farm which doesn't hamper wheat crop sowing and hence doesn't require burning), however the uptake of this policy remains quite low due to high initial investment in the machine (Gupta and Somnathan, 2016). Increasing the subsidy will go a long way in preventing crop residue burning as residue banning has largely remained ineffective and harmful biomass burning continues undeterred.

[16]This is an average figure for South Asia

References

State of Global Air 2018. Technical report, Health Effects Institute.

Vulnerability of forest fire 2017. Technical report, Forest Survey of India.

Acharya, P., Sreekesh, S., and Kulshrestha, U. (2016). Ghg and aerosol emission from fire pixel during crop residue burning under rice and wheat cropping systems in north-west india. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41:753.

Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, 25(3):153–72.

Barber, C. and Schweithelm, J. (2000). Trial by fire: Forest fires and forestry policy in indonesia's era of crisis and reform. Technical report, World Resources Institute, Washington, DC (EUA). Forest Frontiers Initiative.

Bharadwaj, P., Gibson, M., Zivin, J. G., and Neilson, C. (2017). Gray matters: fetal pollution exposure and human capital formation. *Journal of the Association of Environmental and Resource Economists*, 4(2):505–542.

Brainerd, E. and Menon, N. (2014). Seasonal effects of water quality: The hidden costs of the green revolution to infant and child health in India. *Journal of Development Economics*, 107:49–64.

Chakravorty, U., Pelli, M., and Risch, A. (2014). Far away from the forest? fuelwood collection and time allocation in rural india.

Chakravorty, U., Pelli, M., and Risch, A. (2015). Who is buying? fuelwood collection in rural india.

Chay, K. Y. and Greenstone, M. (2003a). Air quality, infant mortality, and the clean air act of 1970. Technical report, National Bureau of Economic Research.

Chay, K. Y. and Greenstone, M. (2003b). The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *The Quarterly Journal of Economics*, 118(3):1121–1167.

Cummins, J., Agarwal, N., Aiyar, A., Bhattacharjee, A., Gunadi, C., Singhania, D., Taylor, M., Wigton-Jones, E., et al. (2017). Month of birth and child height in 40 countries. Technical report.

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

Currie, J. and Vogl, T. (2013). Early-life health and adult circumstance in developing countries. *The Annual Review of Resource Economics*, 5(1):1–36.

Currie, J. and Walker, R. (2011). Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal: Applied Economics*, 3(1):65–90.

Currie, J., Zivin, J. G., Mullins, J., and Neidell, M. (2014). What do we know about short- and long-term effects of early-life exposure to pollution? *The Annual Review of Resource Economics*, 6(1):217–247.

Deaton, A. and Drèze, J. (2009). Food and nutrition in india: facts and interpretations. *Economic and political weekly*, pages 42–65.

Dey, S., Girolamo, L., Van Donkelaar, A., Tripathi, S., Gupta, T., and Mohan, M. (2012). Decadal exposure to fine particulate matters (pm 2.5) in the indian subcontinent using remote sensing data. *Remote Sens Environ*, 127:153–161.

Do, Q.-T., Joshi, S., and Stolper, S. (2018). Can environmental policy reduce infant mortality? evidence from the ganga pollution cases. *Journal of Development Economics*, 133:306–325.

Douglass, R. L. (2008). Quantification of the health impacts associated with fine particulate matter due to wildfires.

Fletcher, J. M., Green, J. C., and Neidell, M. J. (2010). Long term effects of childhood asthma on adult health. *Journal of Health Economics*, 29(3):377–387.

Foster, A., Gutierrez, E., and Kumar, N. (2009). Voluntary compliance, pollution levels, and infant mortality in mexico. *American Economic Review*, 99(2):191–97.

Foster, A. D. and Rosenzweig, M. R. (2003). Economic growth and the rise of forests. *The Quarterly Journal of Economics*, 118(2):601–637.

Frankenberg, E., McKee, D., and Thomas, D. (2005). Health consequences of forest fires in indonesia. *Demography*, 42(1):109–129.

Galasso, E., Wagstaff, A., Naudeau, S., and Shekar, M. (2016). The economic costs of stunting and how to reduce them. *Policy Research Note World Bank, Washington, DC.*

Greenstone, M. and Hanna, R. (2014). Environmental regulations, air and water pollution, and infant mortality in india. *American Economic Review*, 104(10):3038–72.

Gupta, P. K., Sahai, S., Singh, N., Dixit, C., Singh, D., Sharma, C., Tiwari, M., Gupta, R. K., and Garg, S. (2004). Residue burning in rice–wheat cropping system: Causes and implications. *Current science*, pages 1713–1717.

Gupta, R. (2014). Low-hanging fruit in black carbon mitigation: Crop residue burning in south asia. *Climate Change Economics*, 5(04):1450012.

Gupta, S., Agarwal, R., and Mittal, S. K. (2016). Respiratory health concerns in children at some strategic locations from high pm levels during crop residue burning episodes. *Atmospheric environment*, 137:127–134.

Holstius, D. M., Reid, C. E., Jesdale, B. M., and Morello-Frosch, R. (2012). Birth weight following pregnancy during the 2003 southern california wildfires. *Environmental health perspectives*, 120(9):1340.

Isen, A., Rossin-Slater, M., and Walker, W. R. (2017). Every breath you take—every dollar you'll make: The long-term consequences of the clean air act of 1970. *Journal of Political Economy*, 125(3):848–902.

Jain, N., Bhatia, A., and Pathak, H. (2014). Emission of air pollutants from crop residue burning in india. *Aerosol and Air Quality Research*, 14(1):422–430.

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

Kaskaoutis, D., Kumar, S., Sharma, D., Singh, R. P., Kharol, S., Sharma, M., Singh, A., Singh, S., Singh, A., and Singh, D. (2014). Effects of crop residue burning on aerosol properties, plume characteristics, and long-range transport over northern india. *Journal of Geophysical Research: Atmospheres*, 119(9):5424–5444.

Kunii, O., Kanagawa, S., Yajima, I., Hisamatsu, Y., Yamamura, S., Amagai, T., and Ismail, I. T. S. (2002). The 1997 haze disaster in indonesia: its air quality and health effects. *Archives of Environmental Health: An International Journal*, 57(1):16–22.

Neidell, M. J. (2004). Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma. *Journal of health economics*, 23(6):1209–1236.

O'Donnell, M. and Behie, A. (2015). Effects of wildfire disaster exposure on male birth weight in an australian population. *Evolution, medicine, and public health*, 2015(1):344–354.

Rangel, M. A. and Vogl, T. (2016). Agricultural fires and infant health. Technical report, National Bureau of Economic Research.

Rukumnuaykit, P. (2003). Crises and child health outcomes: the impacts of economic and drought/smoke crises on infant mortality and birth weight in indonesia. *Economics Department*.

Sanders, N. J. (2012). What doesn't kill you makes you weaker prenatal pollution exposure and educational outcomes. *Journal of Human Resources*, 47(3):826–850.

Sharma, A. R., Kharol, S. K., Badarinath, K., and Singh, D. (2010). Impact of agriculture crop residue burning on atmospheric aerosol loading–a study over punjab state, india. *Annales Geophysicae (09927689)*, 28(2).

Simeonova, E., Currie, J., Nilsson, P., and Walker, R. (2018). Congestion pricing, air pollution and children's health. Technical report, National Bureau of Economic Research.

Singh, C. P. and Panigrahy, S. (2011). Characterisation of residue burning from agricultural system in india using space based observations. *Journal of the Indian Society of Remote Sensing*, 39(3):423.

Srivastava, P. and Garg, A. (2013). Forest fires in india: Regional and temporal analyses. *Journal of Tropical Forest Science*, pages 228–239.

Vadrevu, K. P., Lasko, K., Giglio, L., and Justice, C. (2015). Vegetation fires, absorbing aerosols and smoke plume characteristics in diverse biomass burning regions of asia. *Environmental Research Letters*, 10(10):105003.

Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., and Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environmental health perspectives*, 118(6):847.

Vijayakumar, K., Safai, P., Devara, P., Rao, S. V. B., and Jayasankar, C. (2016). Effects of agriculture crop residue burning on aerosol properties and long-range transport over northern india: A study using satellite data and model simulations. *Atmospheric Research*, 178:155–163.

Tables and Figures

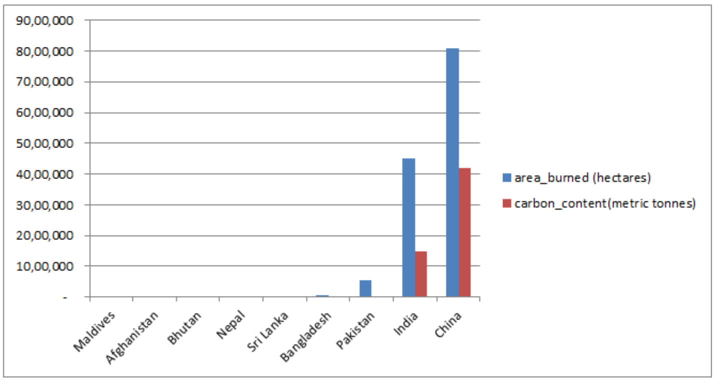

Figure 1: Total area burned and carbon content released due to biomass burning in South Asia and China for year 2015.

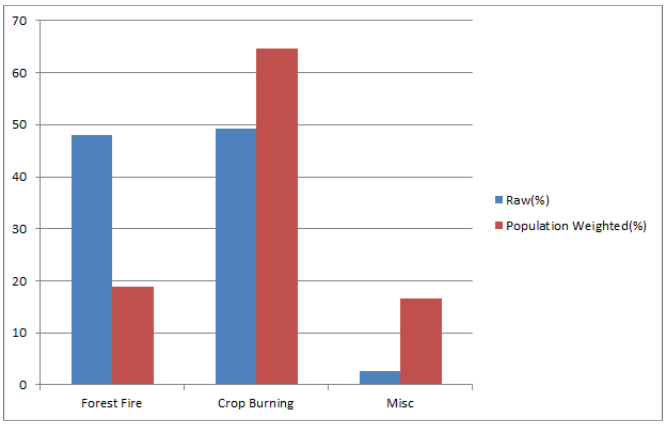

Figure 2: Biomass burning events categorization into forest fires, crop burning events and miscellaneous events. Raw proportion and population weighted proportion for all fire events which took place in India between 2010-2016.

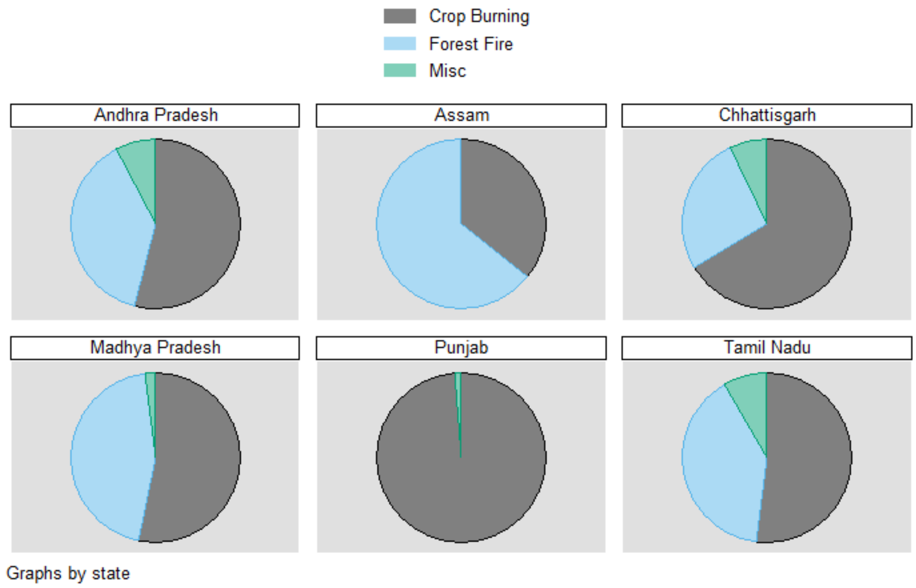

Figure 3: Population weighted split of all biomass burning events which took place from 2010-2016 for select states.

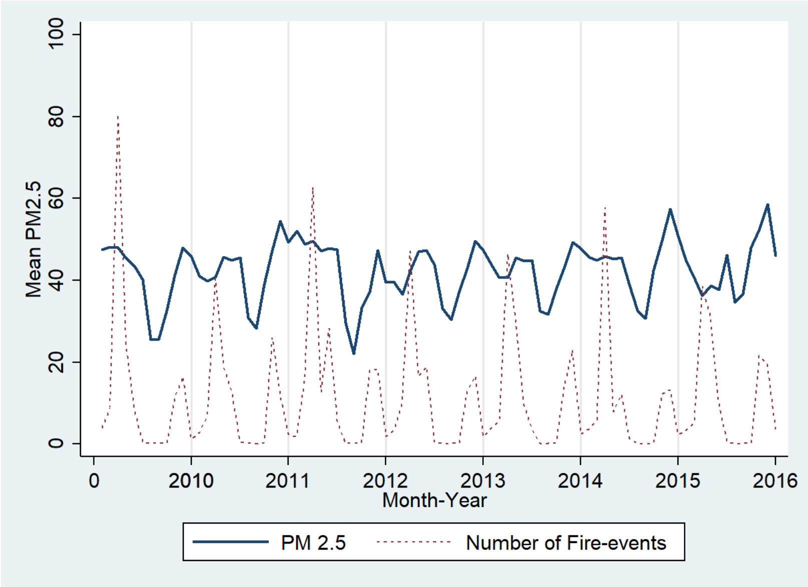

Figure 4: Weighted PM2.5 and Number of fire-events (in 50km radius) for each month in every year from 2010 to 2016. Figure represents mean over all sampled clusters in DHS-4 for India.

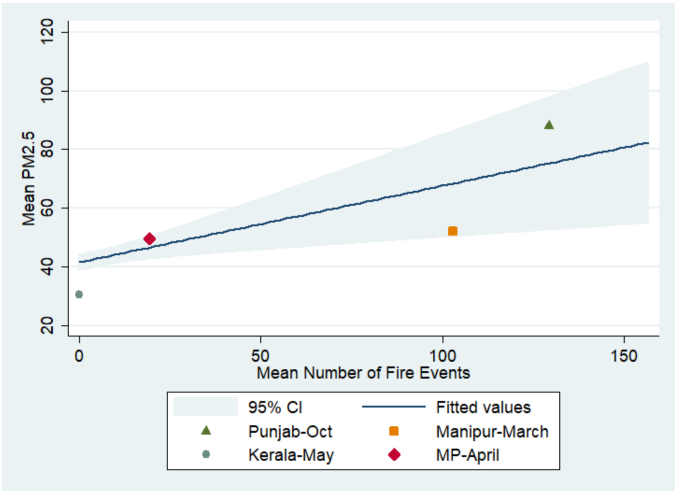

Figure 5: Relationship between mean PM2.5 and mean number of fire-events (in 50 km radius). Unit of observation is state-month, shaded area is 95% confidence interval. Mean values correspond to data from 2010 to 2016.

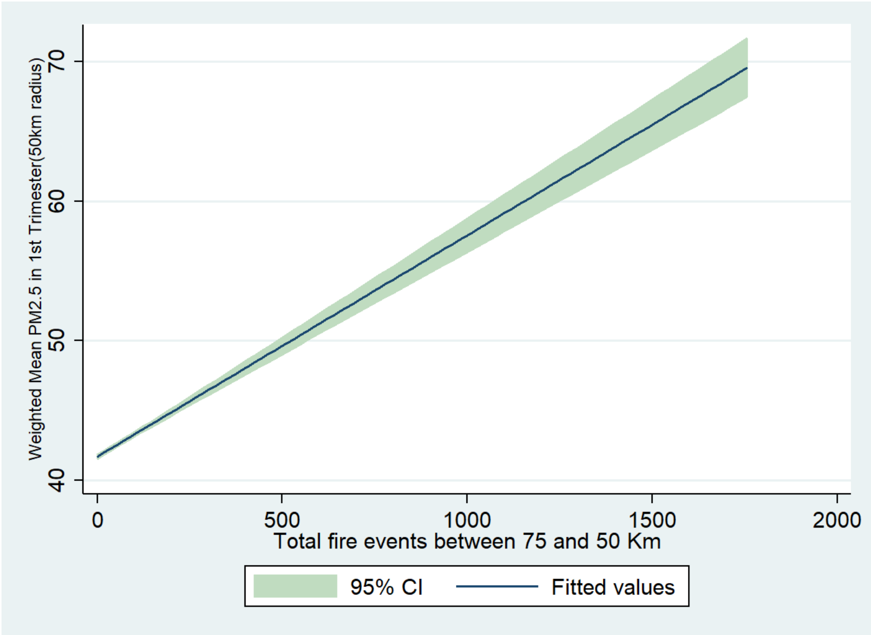

Figure 6: Relationship between mean PM2.5 (in 50 km radius) and Number of fire-events between 75 and 50 km radius(Non-local fire-events). Unit of observation is a child, shaded area is 95% confidence interval.

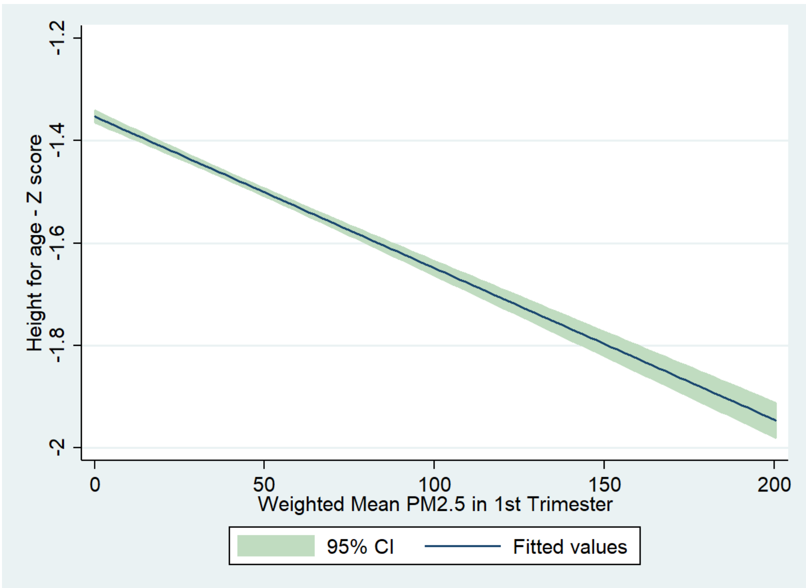

Figure 7: Linear fit plot for relationship between Height-for-age Z scores and weighted mean PM2.5. Unit of observation is a child, shaded area is 95% confidence interval

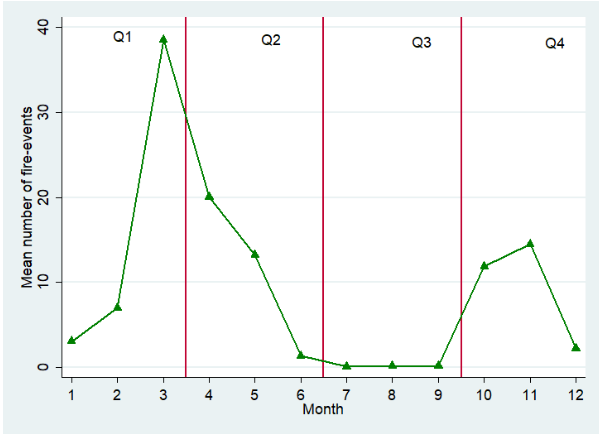

Figure 8: Temporal variation in fire-events: All India mean number of fire-events in 50 km radius for each month. Mean calculated over all sampled clusters from DHS-4. Fire-events data used spans years 2010 to 2016.

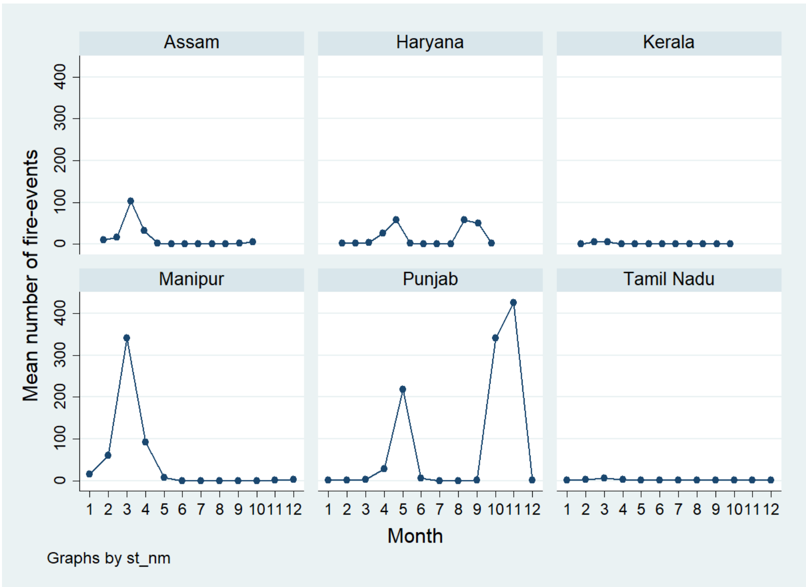

Figure 9: Temporal variation in fire-events (shown here for select states): State level mean number of fire-events in 50 km radius for each month. Mean calculated over all sampled clusters from DHS-4 contained within a state. Fire-events data used spans years 2010 to 2016.

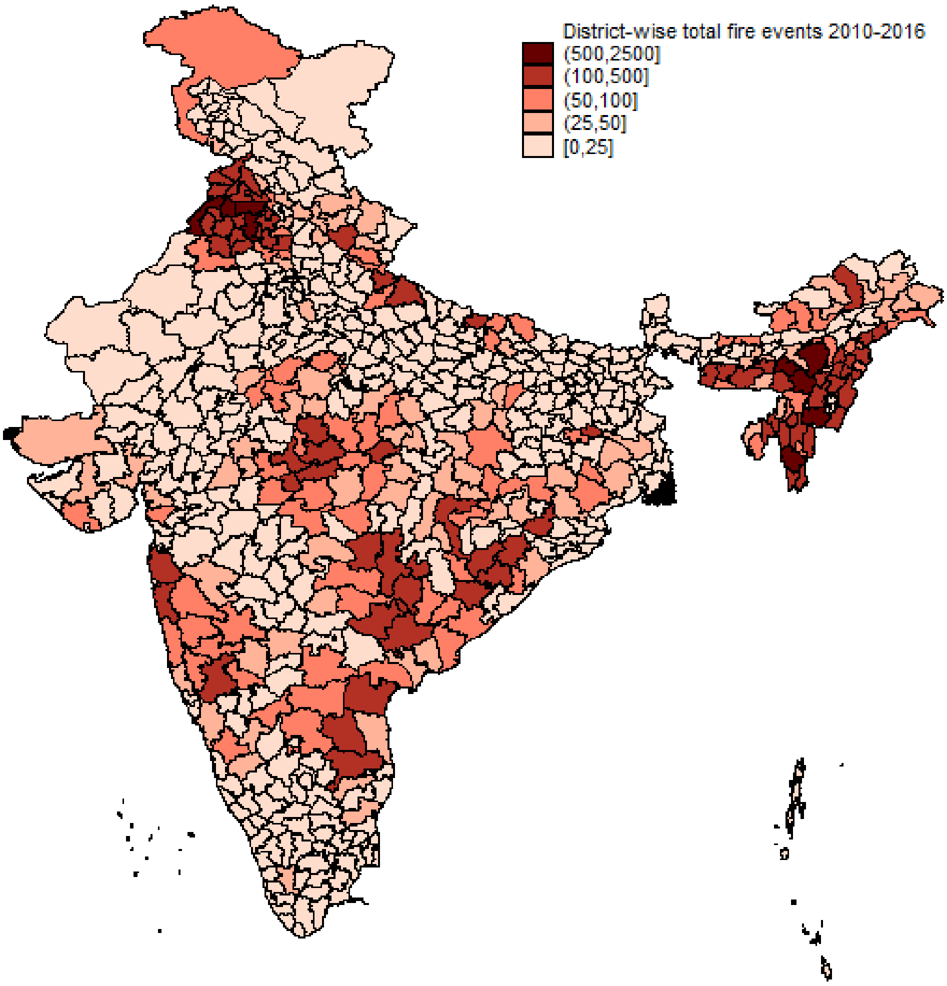

Figure 10: Spatial variation in fire-events: Total fire-events in each district in a year (mean across years 2010-2016).

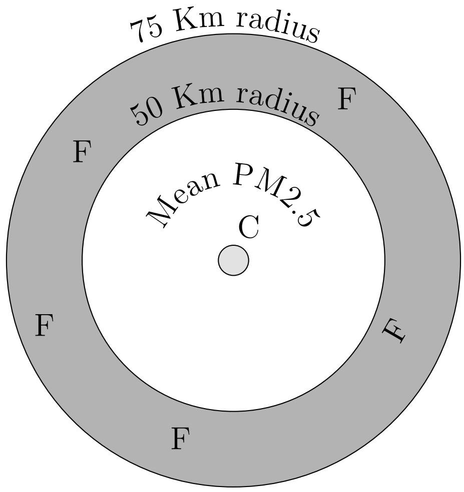

Figure 11: Identification Methodology: Center (smallest grey circle) represents the cluster location, White circle corresponds to 50km radius circle around the cluster location, Grey ring corresponds to area between two circles (75 and 50 Km radii circles) with cluster location as the center. Mean pollution level is calculated for the white circle, we call this *local* pollution level for cluster C. Local pollution level is instrumented using *non-local* biomass burning events which take place in the grey ring.

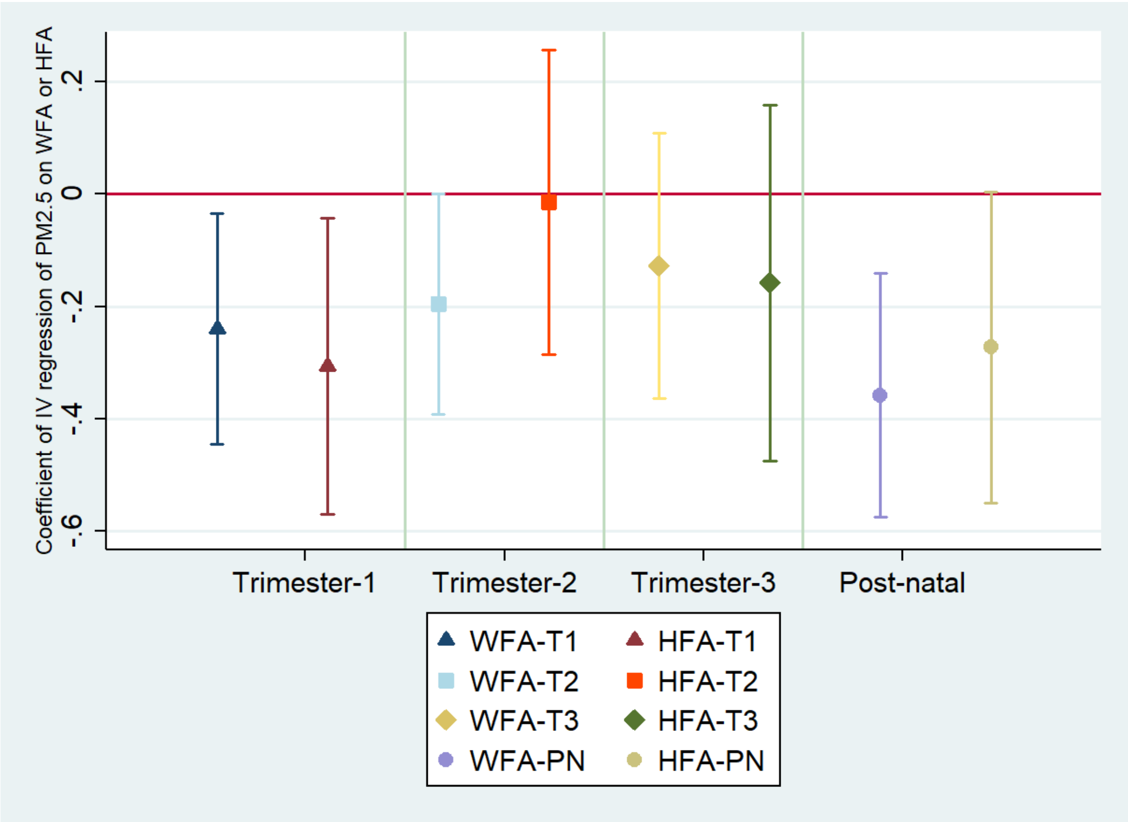

Figure 12: Coefficient of 2SLS regression of outcomes(HFA-Z and WFA-Z) on outdoor air pollution for different critical windows of development of a child. Vertical lines represent 95 % confidence intervals.

Table 1: Summary Statistics

variable	Mean	Std Dev
Outcomes		
Height-for-age Z score	-1.48	1.68
Weight-for-age Z score	-1.53	1.22
Child characteristics		
Dummy for male child	0.52	0.50
Birth order	2.27	1.47
Child's age (in months)	30.00	16.99
Pregnancy duration	9.02	0.47
Parents characteristics		
Mother's age at birth of child	24.54	4.97
Mother's number of education years	6.22	5.13
Dummy for literate father	0.97	0.17
Household characteristics		
Dummy for rural	0.76	0.43
Age of head of household	44.55	15.14
Dummy for head of household being male	0.88	0.33
Dummy for Hindu	0.73	0.45
Household size	6.56	2.86
Owns Fridge	0.24	0.43
Owns TV	0.58	0.49
Owns Car	0.06	0.23
Owns Motorcycle	0.36	0.48
Dummy for electricity connection	0.85	0.36
Dummy for source of water: Pipedwater	0.23	0.42
Dummy for using clean cooking fuel	0.28	0.45
Dummy for open defecation (OD)	0.46	0.50
Fraction of HHs who practice OD in a cluster	0.43	0.35
Pollution measures		
Trimester-1 : Mean weighted PM2.5	42.48	33.63
Trimester-2 : Mean weighted PM2.5	41.20	31.99
Trimester-3 : Mean weighted PM2.5	42.62	33.64
Post-natal : Mean weighted PM2.5	43.06	33.55

Table 2: First-stage regression

	Weighted mean PM2.5 in 50km radius			
	(1)	(2)	(3)	(4)
IV1: Number of fire events between 75 and 50km radius	0.249***	0.379***		
	(0.011)	(0.009)		
IV2: Number of fire events in 50km radius			0.235***	0.358***
			(0.012)	(0.011)
First Stage F-stat	483	1608	366	893
R-square	0.03	0.55	0.03	0.55
Observations	216064	216064	216064	216064
Includes other controls from 2nd stage	Yes	Yes	Yes	Yes
Includes FEs for Quarter, Year, Cluster and their interactions	No	Yes	No	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regressions include controls for gender, birth order and age of child, mother's years of education, mother's age at birth and its square, age and gender of household head, dummy for whether household has pipedwater, has clean cooking source, whether household practices open defecation and fraction of the village who practice open defecation(excluding self). Fire-events variable has been scaled by 10^{-3} and PM2.5 has been scaled by 10^{-2}.

Table 3: Instrumental variable regression of outcomes on weighted PM2.5

	OLS		IV (Fire events between 75 and 50 Km radius)	
	(1) WFA-Z	(2) HFA-Z	(3) WFA-Z	(4) HFA-Z
Trimester-1: Weighted mean PM2.5 in 50km radius	-0.0173	-0.0254	-0.240**	-0.307**
	(0.0171)	(0.0232)	(0.105)	(0.135)
rk LM statistic			1221.58	1221.58
Observations	216064	216064	216064	216064
Mean PM2.5 (unscaled)	42.48			
Includes Child, Mother and Household characteristics	Yes	Yes	Yes	Yes
Includes FEs for Quarter, Year, Cluster and their interactions	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in first trimester. Regressions include controls for gender, birth order and age of child, mother's years of education, mother's age at birth and its square, age and gender of household head, dummy for whether household has pipedwater, has clean cooking source, whether household practices open defecation and fraction of the village who practice open defecation(excluding self). Fire-events variable has been scaled by 10^{-3} and PM2.5 has been scaled by 10^{-2}.

Table 4: Flexible controls for child age: 2SLS regressions of outcomes on weighted PM2.5

	(1) WFA-Z	(2) WFA-Z	(3) WFA-Z	(4) WFA-Z	(5) HFA-Z	(6) HFA-Z	(7) HFA-Z	(8) HFA-Z
Trimester-1: Weighted mean PM2.5 in 50km radius	-0.240**	-0.225**	-0.224**	-0.261**	-0.307**	-0.268**	-0.265**	-0.350***
	(0.105)	(0.105)	(0.105)	(0.105)	(0.135)	(0.135)	(0.135)	(0.135)
Observations	216064	216064	216064	216064	216064	216064	216064	216064
Childage in months (Linear control)	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Childage in months (Quadratic control)	No	Yes	Yes	No	No	Yes	Yes	No
Childage in months (Cubic control)	No	No	Yes	No	No	No	Yes	No
Childage in years (Dummies)	No	No	No	Yes	No	No	No	Yes
Includes Child, Mother and Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Includes FEs for Quarter, Year, Cluster and their interactions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in first trimester. Regressions include controls which are same as the ones mentioned in the notes for table 2. Fire-events have been scaled by 10^{-3} and PM2.5 has been scaled by 10^{-2}.

Table 5: Robustness Check - 1:

i) Including rainfall

ii) Changing "Quarter" definition to quarter as per month at the end of 3 month exposure

	(1) WFA-Z	(2) HFA-Z	(3) WFA-Z	(4) HFA-Z
Trimester-1: Weighted mean PM2.5 in 50km radius	-0.296**	-0.387**	-0.249**	-0.300**
	(0.124)	(0.160)	(0.104)	(0.134)
Mean rainfall	-0.0921**	-0.134***		
	(0.0364)	(0.0472)		
Observations	212349	212349	216064	216064
Includes Child, Mother and Household characteristics	Yes	Yes	Yes	Yes
Fixed Effects:				
Cluster	Yes	Yes	Yes	Yes
Year of conception	Yes	Yes	Yes	Yes
Quarter of conception	Yes	Yes	No	No
Quarter *(as per the end of 3 month exposure)*	No	No	Yes	Yes
Year * Quarter	Yes	Yes	Yes	Yes
Cluster * Year of conception	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in first trimester. Regressions include controls which are same as the ones mentioned in the notes for table 2. Fire-events and rainfall have been scaled by 10^{-3} and PM2.5 has been scaled by 10^{-2}. Rainfall variable corresponds to mean rainfall level in the 50km radius around a cluster.

Table 6: Robustness Check - 2: Alternate radii analysis

	Fire events between 100 and 50 Km		**Fire events between 50 and 30 Km**		**Dropping Punjab**	
	(1) **WFA-Z**	(2) **HFA-Z**	(3) **WFA-Z**	(4) **HFA-Z**	(5) **WFA-Z**	(6) **HFA-Z**
Trimester-1: Weighted mean PM2.5 in 50km radius	-0.215**	-0.271**			-0.314***	-0.394***
	(0.0985)	(0.128)			(0.0974)	(0.133)
Trimester-1: Weighted mean PM2.5 in 30km radius			-0.209*	-0.302**		
			(0.118)	(0.150)		
Observations	216064	216064	209691	209691	211350	211350
Includes Child, Mother and Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Includes FEs for Quarter, Year, Cluster and their interactions	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in first trimester. Regressions include controls which are same as the ones mentioned in the notes for table 2. Fire-events have been scaled by 10^{-3} and PM2.5 has been scaled by 10^{-2}.

Table 7: Robustness Checks - 3

	Mother's Education	Father Literate	Asset Ownership	Poor	Polio Vaccination	HH size	Pregnancy Duration
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Number of fire events between 75 and 50km radius in 1st trimester	-0.0358	0.00701	0.0178	-0.00808	0.0266*	0.0198	0.0121
	(0.138)	(0.00460)	(0.0320)	(0.0113)	(0.0140)	(0.0662)	(0.0142)
Observations	223770	223770	223770	223770	223770	223770	223770
Includes Child, Mother and Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Includes FEs for Quarter, Year, Cluster and their interactions	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regressions include controls for meduyrs feducation headage headmale wealthix. Fire-events have been scaled by 10^{-3}.

Table 8: Do mothers plan conception?

	(1)	(2)
	Number of conceptions each month	
Mean number of fire events in 50 km radius	-0.000255	
	(0.00310)	
Mean number of fire events between 75 and 50km radius		0.00462
		(0.00294)
Observations	22,89,564	22,89,564
Cluster FE	Yes	Yes
Month FE	Yes	Yes

Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regressions include controls for mother's years of education, literacy status of father, age and gender of household head and wealth index of household. Fire-events have been scaled by 10^{-3}. Unit of analysis is mother-month, dependent variable is number of conceptions that a mother has in each month for all births that happened in last 5 years.

Table 9: Heterogeneity

	2SLS regression: Height-for-age Z score			
	Rural	**Urban**	**North**	**South**
	(1)	(2)	(3)	(4)
Trimester-1: Weighted mean	-0.287*	-0.485*	-0.307**	-0.842
PM2.5 in 50km radius	(0.156)	(0.262)	(0.129)	(1.040)
Observations	164300	51764	164481	51583
Includes Child, Mother and Household characteristics	Yes	Yes	Yes	Yes
Includes FEs for Quarter, Year, Cluster and their interactions	Yes	Yes	Yes	Yes

Each coefficient corresponds to an individual regression of HFA-Z on on weighted mean PM2.5 in first trimester. Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Regressions include controls which are same as the ones which are mentioned in the notes for table 2. Fire-events have been scaled by 10^{-3}.

North Indian states: Arunachal Pradesh, Assam, Bihar, Chandigarh, Gujarat, Haryana, Himachal Pradesh, Jammu and Kashmir, Jharkhand, Madhya Pradesh, Manipur, Meghalaya, Mizoram Nagaland, Delhi, Punjab, Rajasthan, Sikkim, Tripura, Uttar Pradesh and Uttarakhand.

South Indian states: Andhra Pradesh, Karnataka, Kerala, Maharashtra, Chhattisgarh, Odisha, Telangana, West Bengal, Lakshwadeep Islands, Andaman and Nicobar Islands, Dadar and Nagar Haveli, Daman and Diu, Puducherry and Goa.

Table 10: Instrumental variable effects:

Impact of weighted PM2.5 in 2nd Trimester to Post-natal period

	IV	
	(1)	(2)
	WFA-Z	HFA-Z
Trimester-2: Weighted meanPM2.5 in 50km radius	-0.196*	-0.0144
	(0.100)	(0.138)
Observations	219847	219847
Trimester-3: Weighted meanPM2.5 in 50km radius	-0.127	-0.157
	(0.121)	(0.161)
Observations	219844	219844
Post-natal: First 3 months after birth	-0.357***	-0.272*
Weighted mean PM2.5 in 50km radius	(0.111)	(0.141)
Observations	213116	213116
Includes Child, Mother and Household characteristics	Yes	Yes
Includes FEs for Quarter, Year, Cluster and their interactions	Yes	Yes

Standard errors in parentheses are clustered by DHS cluster. Notation for p-values *** is $p < 0.01$, ** is $p < 0.05$ & * is $p < 0.1$. Each coefficient corresponds to an individual 2SLS regression of HFA-Z or WFA-Z on weighted mean PM2.5 in different time windows. Regressions include controls which are same as the ones mentioned in the notes for table 2. Fire-events variable has been scaled by 10^{-3} and PM2.5 has been scaled by 10^{-2}.