Air Pollution and Child Health in Urban India

Arkadipta Ghosh and Arnab Mukherji

Arkadipta Ghosh
Mathematica Policy Research
Princeton, NJ
arkadipta@gmail.com

Arnab Mukherji
Center for Public Policy,
IIM Bangalore
arnab.mukherji@gmail.com

Abstract

A potential source of confounding in studies investigating the effect of indoor air pollution on child health is exposure to ambient air pollution. We investigate this relationship pairing city-level air pollution measures with child level data from the National Family Health Survey (2005-06) for six cities in India. We address simultaneity in child health outcomes and potential endogeneity of city-level air pollution by using a bivariate probit regression framework with city fixed effects. Our findings show –1) an increase in ambient air pollution significantly increases child morbidity; 2) the type of cooking fuel used at home (usual measure of indoor pollution) is not a significant determinant of child morbidity once ambient air pollution and other child, household, and city-level covariates are controlled for; and 3) it is important to explicitly account for the correlation in various child health outcomes by modeling them jointly. Our findings suggest that targeted city-wide reductions in ambient air pollution could play an important role in improving child health.

Word Count
Abstract: 163 words
Paper: 1893 words

Keywords
Ambient Air Pollution, Child Health, Bivariate Probit, Fixed Effects, NFHS

1. **Introduction**:

Exposure to air pollution has been linked to poor child health outcomes in a range of studies that have looked at a variety of health measures.[1] For example, Frankenberg et al. (2004) investigate the effect of outdoor air pollution due to the forest fires in Indonesia in 1997 on infants and reports that there was a 1% decline in the Indonesian cohort size due to these fires; (Smith 2000) on the other hand looks at the impact of solid fuels used for cooking at home and suggests that as much as 4-5% of the national disease burden for India may be explained by indoor air pollution alone. A limitation in the current literature is the lack of objective measures of both indoor and outdoor air pollution in the same analysis. Additionally, several studies use air pollution proxies such as the occurrence of forest fires (Jayachandran 2009) or the type of cooking fuels used at home (Smith 2000), in the absence of more direct pollution measures. An alternative to using such proxies is to use air pollution data gathered from direct observation with expensive monitoring instruments.

In this paper, we combine directly measured ambient air pollution data for 2005-06, collected by the National Air Monitoring Program of the Ministry of Environment and Forests in India with household and child-level data from the third wave of the National Family Health Survey or NFHS-3 (2005-6). This allows us to construct an analysis sample that not only has variation in the type of solid fuel used at home (our proxy for indoor air pollution, as in some of the previous studies), but also variation in the average level of air pollution that households are exposed to during the month of their interview. Hence, we are able to investigate the relative effects of both types of exposure to air pollution on child morbidity, as captured by the incidence of two common illnesses in children – cough and fever – in the week prior to the interview.

Apart from using both ambient and indoor pollution measures, we also explicitly tackle concerns on model specification and causality in this paper. With repeated observations from the same city, we are able to control for unobserved city fixed effects that may uniformly affect all children in the same city, for example, latitude, inherently

[1] See Bruce et. al. (2000), Chay and Greenstone (2003), Frankenberg et. al. (2004), Currie el. al. (2005) & Janke et. al. (2009).

high versus low levels of vehicular emission, location near an industrial hub, proximity to a river, etc. We also address potential misspecification concerns by jointly modeling the probability of a child having a fever and that of having a cough, since they both pertain to the same underlying health status of the child and are likely to be determined by similar factors – both internal and external to the child. This is an important departure from the literature where these are usually studied as independent events.

We find that a rise in ambient air pollution significantly increases the likelihood of a child suffering from cough and fever in the past week. However, the type of cooking fuel used at home is not significantly related to child morbidity after accounting for ambient air pollution and other child- and household-level control variables. Thus, while bad air is bad for child health, we find that ambient air pollution is a more significant determinant of the child health outcomes we study. This suggests that controlling city-wide air pollution could significantly lower child morbidity, and should receive greater emphasis in urban planning and infrastructure development. We also find a significant correlation between the two child morbidity outcomes – fever and cough, which suggests that models that do not explicitly account for this correlation are likely to be mis-specified.

The rest of the paper is arranged as follows: Section 2 presents the background and the data we use, Section 3 presents our empirical strategy, Section 4 presents our results, and Section 5 closes with a discussion of our findings.

2. Background and Data

That poor air quality leads to poor health outcomes for both adults and children is fairly well established in the literature. The mechanisms by which air quality affects health is usually thought to be through reduced pulmonary functioning leading to acute respiratory symptoms (Bruce et. al. 2000). To the best of our knowledge, the focus in this literature has been on either ambient air pollution or on indoor air pollution, but not both. One of our contributions in this paper lies in examining the relative effects of both sorts of exposure on child health.

We use data from the city sample of NFHS wave 3 that was collected over 2005 and 2006. For our analysis, we specifically use the child-recode datafile (IAKR51FL.DTA), and the health outcomes analyzed are incidence of fever and cough among children (born within the last three years) during the two weeks prior to the interview, as reported by the household respondent. Data on ambient air pollution is taken from the administrative records maintained by the Central Pollution Control Board (CPCB), Government of India, under its National Air Monitoring Program (NAMP). NAMP provides data on four key pollutants for the cities on which we have child-level data over the survey duration 2005-6[2]: Sulphur Dioxide (S02), Nitrogen Dioxide (NO2), Respirable Suspended Particulate Matter (RSPM / $PM_{2.5}$), and Suspended Particulate Matter (SPM/PM_{10}). We use the average monthly levels of RSPM and SPM as our primary measures of ambient air pollution in separate specifications, since 1) particle pollution that consists of microscopic solids or liquid droplets can affect the lungs and cause health problems including bronchial irritation, coughing, decreased lung function, aggravated asthma, and chronic bronchitis, 2) fine particles or RSPM are 2.5 micrometers in diameter or smaller and are more likely to be found in smoke and haze – ubiquitous features of most major Indian cities – thus constituting a major threat to respiratory health and functioning; 3) coarse particles or SPM that are larger than 2.5 micrometers but smaller than 10 micrometers in diameter, and more common on roadways and in dust, also pose significant health problems; 3) the high correlation (> 0.95) between RSPM, SPM, NO2 precludes their joint inclusion in the same specification; and 4) the average monthly levels of SO2 and NO2 were usually within their permissible levels for most cities in our sample. Figures 1 - 4 plot the city-level monthly average, maximum, minimum, and NAMP-stipulated permissible level of each of the four pollutants on which data are available from the NAMP across the NFHS interview months.[3] We combine air pollution data with the NFHS dataset by calculating standardized monthly averages or deviations from permissible levels for each source of ambient air pollution and pairing it to each case's month of interview as reported in the NFHS. Data on indoor air pollution comes from the NFHS where we capture each household's indoor air quality using the type of cooking fuel used by the household. Cooking fuel is believed to be the most important

[2] CPCB is a statutory body under the Ministry of Environment and Forests (MoEF). CPCB's primary responsibilities include the prevention, control and abatement of air and water pollution in India.
[3] We check for the effect of S02 and NO2 emissions on child health by estimating additional regressions that we discuss later; our main focus remains the effects of RSPM and SPM on child morbidity.

source of indoor air pollution. From the NFHS data, we know if households use electricity, LPG, natural gas, kerosene, coal, lignite, charcoal, wood, straw/ shrubs/grass, crop residue, or animal dung as the primary cooking fuel. We classify these into three categories, clean cooking fuel (i.e. electricity, LPG, natural gas), unclean fuel (i.e. kerosene, coal, lignite and charcoal), and unprocessed fuel (wood, straw, crop residue, and animal dung).

3. Estimation Strategy

Let child *i*, living in city *c*, in month *m* have a latent propensity to have fever and cough be captured by $\boldsymbol{h}^*_{\boldsymbol{imc}} = (h^*_{1imc}, h^*_{2imc})$ where the $h*$ represents unobserved latent propensities that are only partially observed. These latent propensities are related to a number of child-specific, household-specific, month-specific and city-specific effects; in these models we are specifically interested in the month- and city-specific ambient air pollution variables. Thus we have:

$$\begin{aligned} &h^*_{1imc} = \beta_1(inAP)_{imc} + \beta_2 f(outAP)_{mc} + \boldsymbol{x}_{ic}\boldsymbol{\beta} + \tau_c + \varepsilon_{1imc};\\ &h_{1imc} = 1 \text{ if } h^*_{1imc} > 0 \qquad (1)\\ &h_{1imc} = 0 \text{ otherwise}\end{aligned}$$

$$\begin{aligned} &h^*_{2imc} = \gamma_1(inAP)_{imc} + \gamma_2 f(outAP)_{mc} + \boldsymbol{x}_{ic}\boldsymbol{\gamma} + \tau_c + \varepsilon_{2imc};\\ &h_{2imc} = 1 \text{ if } h^*_{2imc} > 0 \qquad (2)\\ &h_{2imc} = 0 \text{ otherwise}\end{aligned}$$

However, neither h^*_{1imc} and h^*_{2imc} are observed; we only know if they are positive, and thus, we observe (h_{1imc}, h_{2imc}). If we assume that ε_{1imc} and ε_{2imc} are independently distributed N(0,1) then we can estimate two sets of independent probit regressions to estimate the underlying regression coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. However, this essentially assumes that these two health conditions are not correlated – i.e. the child's likelihood of having a fever is unrelated to having a cough. If the two are in fact correlated, as they are likely to be through unobserved child, household, or environmental attributes, then a bivariate probit model that assumes that the errors in these two regressions are jointly distributed is more appropriate, i.e.

$$\varepsilon_{1imc}, \varepsilon_{2imc} \sim N(0,0,1,1,\theta) \qquad (1)$$

where θ is the correlation coefficient between the two error terms. This framework, therefore, allows us to explicitly account for the likely correlation between the two child morbidity measures that are affected by the level of air pollution. In equations (1) and (2),

$inAP_{imc}$ are a set of dummy variables for the use of coarse bio-fuels for cooking at home as a proxy for indoor air pollution, while $outAP_{mc}$ denotes a specific outdoor air pollution measure such as deviation of the monthly average of RSPM or SPM from its permissible level.[4] Apart from these measures, we also account for a number of child-specific variables such as sex, age, and health endowments (height and weight); as well as parental or household variables such as religion, education of the parent, social group, and wealth of the household. The τ_c are city-level fixed effects that account for city-specific unobserved attributes that could influence a child's health as well as ambient air pollution but do not vary either across households within the same city or over time (for example, high versus low volume of traffic, latitude or geographical proximity to say hills, forests, rivers, or to significant sources of pollution, such as a manufacturing hub).

4. Results

Child health is known to be quite fragile and particularly so in developing countries. Even in our data, respondents report a high frequency of child morbidity in terms of the incidence of fever and cough that occurred in the two weeks prior to the NFHS interview (see Figure 2). The out-door air pollution measures tend to be highly correlated indicating months in which SO2 is high is also like to be a month with high NO2, RSPM and SPM (see Table 2). As mentioned earlier, there is a large amount of variation across months, and across cities in the level of outdoor air pollution that is reported, and in the case of many months, measured pollutants tend to be well below the levels mandated in Table 1. SPM and RSPM tend to be the pollutants that most frequently violate these safe limits and our main focus is therefore on particle pollution in this analysis. In addition, our sample consist of children whose average age is about 29 months and for these children we observe their gender, native health status as measured by height for age and weight for height, their household asset status, social group membership, religion, mother's education etc (see Table 3).

We start by estimating a bivariate probit specification for cough and fever that only includes the indoor air pollution variables and a time trend. Next, we progressively

[4] We use a number of alternative functional forms of these biweekly readings for our analysis – the mean, standardized mean, the standard deviation calculated over the month, the coefficient of variation, and finally, deviations in the monthly average levels from the permissible or safe levels as defined by NAPM.

add more covariates to this model – starting with our outdoor air pollution measures, city fixed effects, and finally adding child- and household-level covariates as well. Our estimates suggest that there is a substantial correlation between the two health outcomes that is important to account for in the model. Second, in the basic model with only indoor air pollution measures and monthly time trend, we find that cooking at home with dirty fuels (like coal or lignite) or with unprocessed fuels (like grass, dung, straw, etc.) that generate smoke in comparison to LPG, or electricity based cooking devices tend to lead to a greater likelihood for a child having a cough in the past week. There appears to be no similar effect on fever, although the correlation coefficient across the joint models for having a fever and a cough are statistically significant and positive suggesting that unobserved factors that cause a fever are strongly related to unobserved factors that cause a cough as well. The correlation coefficient is about 0.83 across all the models in Table 4 suggesting that this correlation is sizeable. In Model (2) we find that the indoor air-pollution coefficients are reduced in magnitude and statistical significance on introducing one of our measures of outdoor air pollution - the log of RSPM to Model (1). While indoor air-pollution remains significant, the log of RSPM also has a positive and statistically significant effect on the incidence of cough, suggesting that both forms of air-pollution may be important.

In Model (3) we introduce city level fixed effects to capture variation in unobservables that affect child health across cities but are constant within the city – for example, elevation, humidity, population density etc. On introducing city fixed effects we find that indoor air pollution measures have marginally smaller coefficients, and the coefficient on unprocessed fuels is no longer significant while that on dirty fuels remains significant but is smaller. However, on introducing city fixed effects the coefficient on log of RSPM is about 5 times larger and is significant across both the cough and fever equations suggesting that earlier estimates may be biased by city level unobservables. Note also that outdoor air pollution matters for both fever and cough. Model (4) is our final model in Table 4 where apart from city level fixed effects we also include a number of child, mother, and household level variables that tend to all vary within the city and so captures additional degrees of variation. Once we include individual and household level covariates in addition to city level fixed effects we find that indoor air pollution proxies –

the type of cooking fuel used at home is not statistically significant. The log of RSPM, our measure of outdoor or ambient air pollution, remains positive and statistically significant and is somewhat larger than in Model (3) for both fever and cough.

In Table 5 we explore other measures of outdoor air pollution such as SO2, NO2 and SPM. In this table we estimate the equivalent specification of Model (4) with all its covariates, fixed effects, and indoor air pollution measures, but with different outdoor pollution measures. Log of SO2 has coefficients that are larger than what we see for RSPM, while the NO2 coefficients are not statistically significant. Finally, SPM has coefficients that are also larger than what we see for RSPM. Thus, we see that there is a fair amount of heterogeneity across different air pollution measures in how they affect child health. Apart from the innate differences across different measures that may explain these differences, Figures 1 – 4& 2 also show that there is wide variation in the prevalence of these measures. A natural question that emerges is what happens to child health when the level of outdoor pollution exceeds prescribed norms in Table 1.

In Table 6 we look at the effect of deviating from National Ambient Air Quality Standards (NAAQS) recommended safe levels of outdoor air pollution. Negative deviations indicate air pollution levels are safe while positive deviations indicate that the air pollution levels are harmful or above the permissible level. For ease of interpretation we scale the deviation by diving by hundred and retain the same set of covariates as in Models (4)-(7). Model (8) looks at the effect of deviations in RSPM on a child's fever or cough and finds that the probability of having a fever or a cough is significantly and positively correlated with deviation from the safe levels. Model (9) looks at the effect of deviations in SPM levels and finds a similar relationship between deviations in SPM from its safe level and the likelihood of having a cough or a fever, although size of the coefficients are smaller. Thus, while RSPM and SPM are closely correlated in their variation over time and space it would appear that they influence child health somewhat differently.

So far, in these regression models we have reported estimates for our regression coefficients and not the marginal effects that are computationally more complicated in a limited dependent variable framework. Therefore, for ease of interpretation, we plot the predicted probabilities of the joint outcomes against different levels of outdoor air pollution. Figure 3 plots the joint probability of having both a cough and fever, of having a cough alone, and of having a fever alone, against different levels of the prevalence of RSPM measured on a log scale. We first note that the standard errors are tightly estimated that allows us to conclude that the probabilities of each of these separate events are statistically different from each other. Secondly, the joint probability of having both a fever and a cough is always greater than the probability of having either a fever alone or a cough alone, suggesting that a joint model is the relevant framework for this analysis. Also, almost always, the probability of having a cough alone is more than having a fever alone, except for low levels of RSPM. In fact, the probability of a fever alone declines almost asymptotically with increasing RSPM suggesting that with increasing RSPM, and aggravated respiratory problems, other complications (possibly due to infections) may also set in so that having both a fever and a cough are the most likely state for a child at high pollution levels. Figure 4 and Figure 5 look at the predicted probabilities in relationship to safe levels of pollution for RSPM and SPM respectively. We find the distribution of predicted probabilities to be similar to what we see in Figure 3.

In Table 7 we carry out a simple simulation exercise to get a sense of the marginal effects at the mean. For each pollution measure, we estimate the sample-wide average joint probability of fever and cough when a specific outdoor air pollution variable is held at the sample mean and then again when it is held at the sample maximum. We do these calculations for log of RSPM, scaled deviation of RSPM from safe levels and scaled deviation of SPM from safe levels. The log of RSPM calculations show that the joint probability of having both a cough and a fever goes up as much as three times in going from the mean to the maximum, and similarly the probability of having cough alone goes up from 0.07 to 0.16 on the probability scale also suggesting large effects. Having only fever is a rare event and the point estimate actually declines when going from the mean to the maximum, which we interpret as being indicative of the fact that at high pollution levels a child is increasingly likely to have a cough and possibly also a fever but not just fever. The effects of moving from mean deviation to maximum deviation in RSPM are similar, but the levels are different in comparison to the calculations involving the log of RSPM reflecting the difference in scale between the two measures. The last set of estimates are for deviation in SPM and we again see the same pattern of there being little change in the probability of having fever alone, but that of having both fever and cough, or cough alone increases significantly.

5. Conclusions

Simple omitted variable bias has been known to confound many estimates and here we present evidence to suggest that this may also be the case for the literature investigating the effect of indoor air pollution on child health without explicitly controlling for outdoor air pollution. If higher levels of ambient air pollution lead to worse health outcomes and are also correlated with the levels of indoor air pollution, then the estimated effect of indoor air quality on child health is likely to be an overestimate. Our findings suggest that greater emphasis needs to be placed on improving ambient air quality in general as part of urban planning and development, and policies targeted at city-wide reductions in air pollution can significantly improve child health.

As noted by Kjellström et al. (2006), several interventions have been shown to be cost effective in controlling air pollution in the context of the United States in that the cost of implementing each of these interventions was less than the value of lives saved. These include interventions such as controlling coal-fired power plant emissions through high chimneys and other means, reducing lead in gasoline from 1.1 to 0.1 grams per gallon, and controlling SO2 emission by desulfuring residual fuel oil. Also, they rightly note that this list of cost-effective interventions for controlling air pollution could increasingly become relevant for developing countries as their industrial and transportation pollution situations become similar to that in the United States a few decades back. Larssen et al. (1997) evaluate the cost effectiveness of several measures to control air pollution in Mumbai and the Greater Mumbai area, and recommend that the following measures should be prioritized—inspection and maintenance of vehicles, introduction of unleaded gasoline, and introduction of low-smoke lubricating oil. They also note that controlling the resuspension of road dust would be one of the most cost effective ways of reducing exposure to suspended particles.

While most Indian cities have taken certain steps to control ambient air pollution, e..g., the introduction of compressed natural gas (CNG) for auto-rickshaws in Delhi, phasing out of older vehicles that do not comply with emission norms across cities, and placing greater emphasis on mass transit, pollution levels especially the level of suspended particles—a significant threat to health—continue to be dangerously high in many cities. The findings in this paper clearly show that ambient air pollution significantly increases the risk of respiratory ailments in children, and therefore pose a risk to their future health and well-being as well, since certain respiratory illnesses in childhood could turn into chronic conditions that last a lifetime and significantly hamper an individual's quality of life. These findings, therefore, should reinforce the sense of urgency with which central, state, and local agencies need to move with respect to taking steps to enforce existing regulations and designing creative strategies to control pollution in the fast growing and large urban centers of India.

References

Bruce, N., Perez-Padilla, R. and Rachel Albalak (2000). Indoor air pollution in developing countries: a major environmental and public health challenge. *Bulletin of the World Health Organization*. 78(9), 1078-1092

Currie J. and M. Neidell. (2005). Air Pollution and Infant Health: What can we learn from California's Recent Experience? *Quarterly Journal of Economics*. 120(3), 1003-1030

Chay K and M. Greenstone. (2003). The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," *Quarterly Journal of Economics*, 118(3), 1121–1167.

Duflo,E, M. G. and R. Hanna (2008). Indoor air pollution, health and economic well-being. *Surveys and Perspectives Integrating Environment and Society 1.1*(1).

Frankenberg, E., D. McKee, and D. Thomas (2004): "Health Consequences of Forest Fires in Indonesia," *Demography*, 42(1), 109–129.

Janke, K., Propper, C. and John Henderson. (2009). Do current levels of air pollution kill? The impact of air pollution on population mortality in England. *Health Economics* 18(9), 1031-1055

Jayachandran, S. (2009). Air quality and early-life mortality during Indonesia's massive wildfires in 1997. *Journal of Human Resources 44*(4), 916–954.

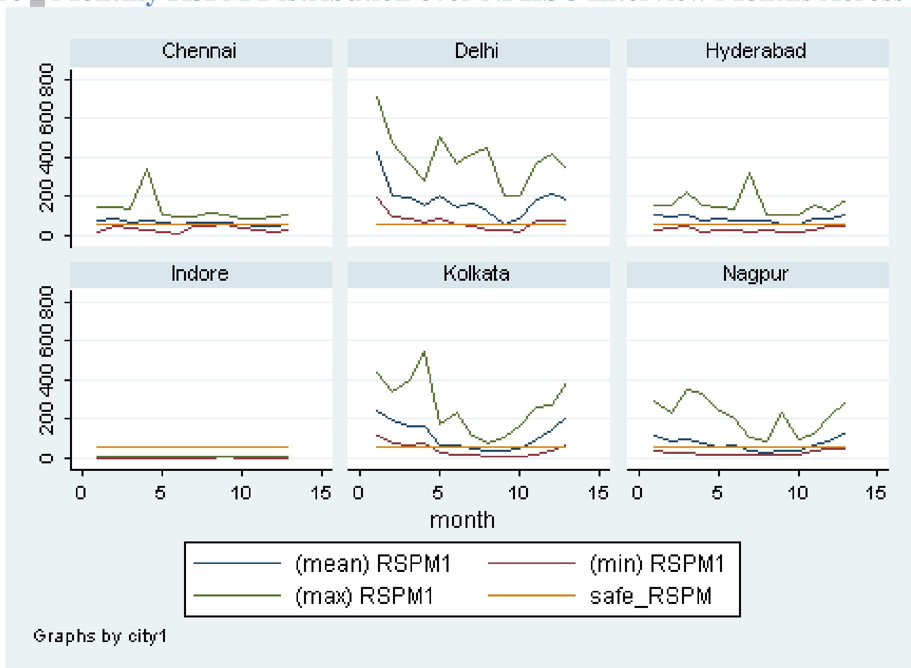
Kjellström, Tord, Madhumita Lodh, Tony McMichael, Geetha Ranmuthugala, Rupendra Shrestha, and Sally Kingsland, "Air and Water Pollution: Burden and Strategies for Control." 2006. Disease Control Priorities in Developing Countries (2nd Edition),ed. , 817-832. New York: Oxford University Press.

Larssen, S., F. Gram, L. O. Hagen, H. Jansen, X. Olsthoorn, K.H. Mehta, R. V. Aundhe, U. Joglekar, and A.A. Mahashur. 1997. URBAIR Urban Air Quality Management Strategy in Asia: Greater Mumbai Report (eds. J.J. Shah and T. Nagpal). World Bank Technical Papers no. 381. Washington, DC: World Bank.

Smith, K. R. (2000). National burden of disease in India from indoor air pollution. *Proceedings of the National Academy of Science 97*(24), 13286–13293.

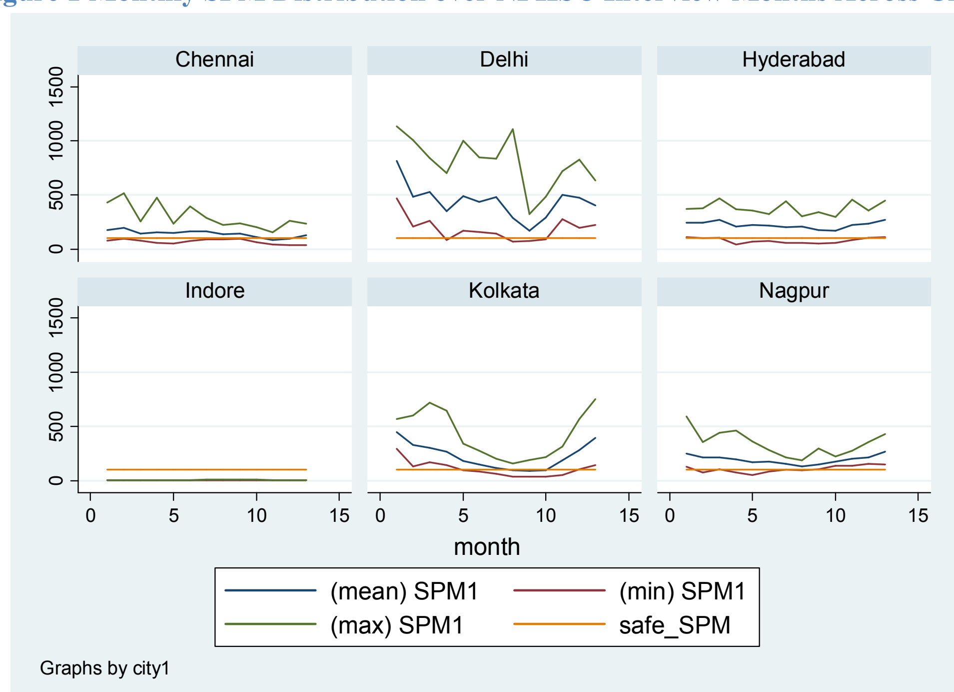
Figures

Figure 1 Monthly RSPM Distribution over NFHS 3 Interview Months Across Cities



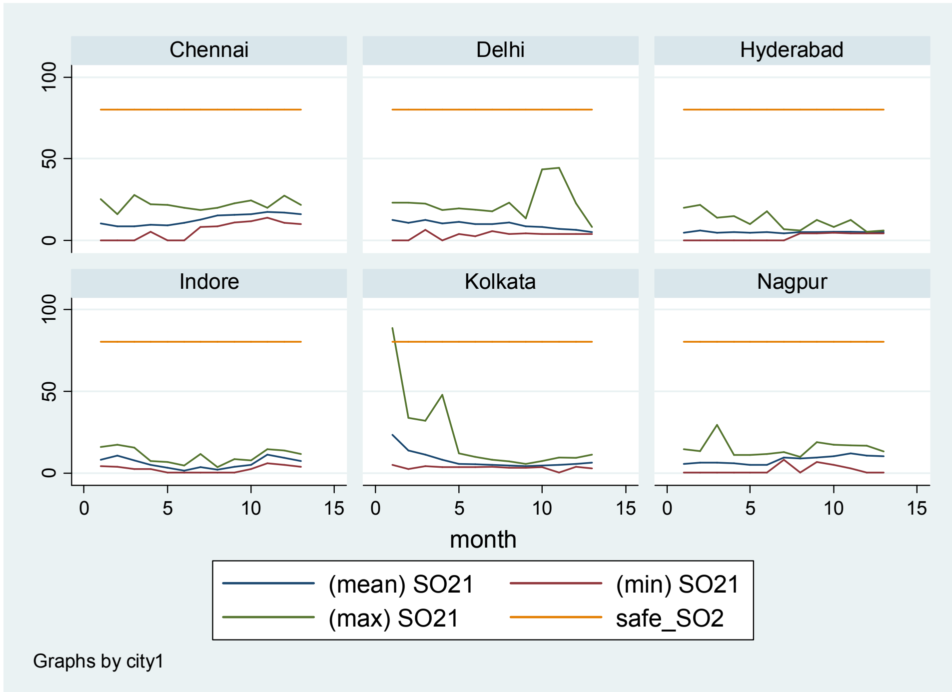
Source: Environmental Data Bank, MOEF, Govt. of India. Safe_RSPM is the level of RSPM that is advised as being safe. Each pollutant is measured in μm^3.

Figure 1 Monthly SPM Distribution over NFHS 3 Interview Months Across Cities



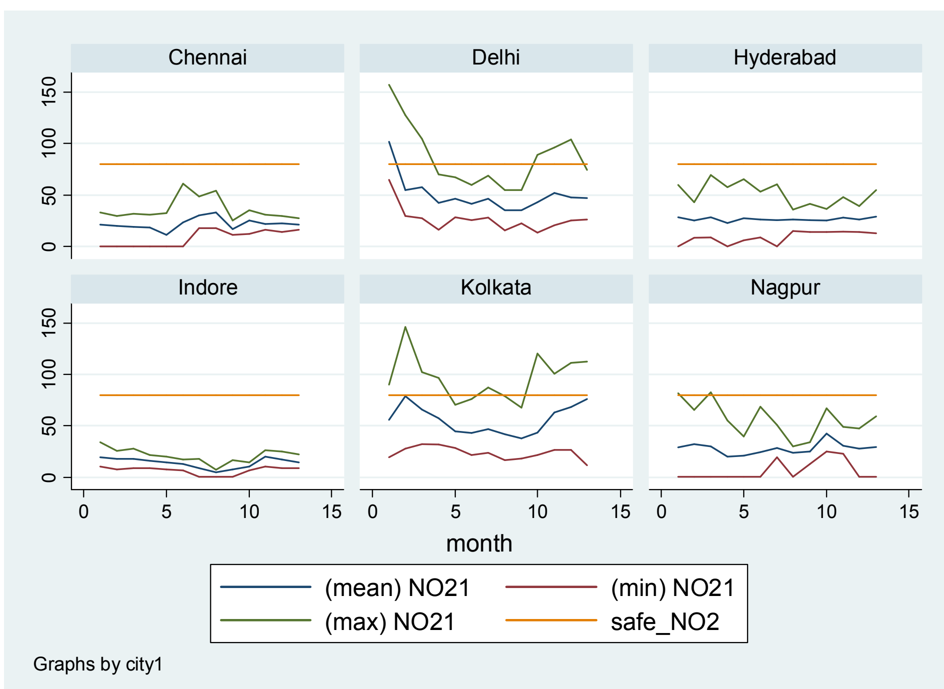
Source: Environmental Data Bank, MOEF, Govt. of India. Safe_SPM is the level of SPM that is advised as being safe. Each pollutant is measured in μm^3.

Figure 3 Monthly SO2 Distribution over NFHS 3 Interview Months across cities



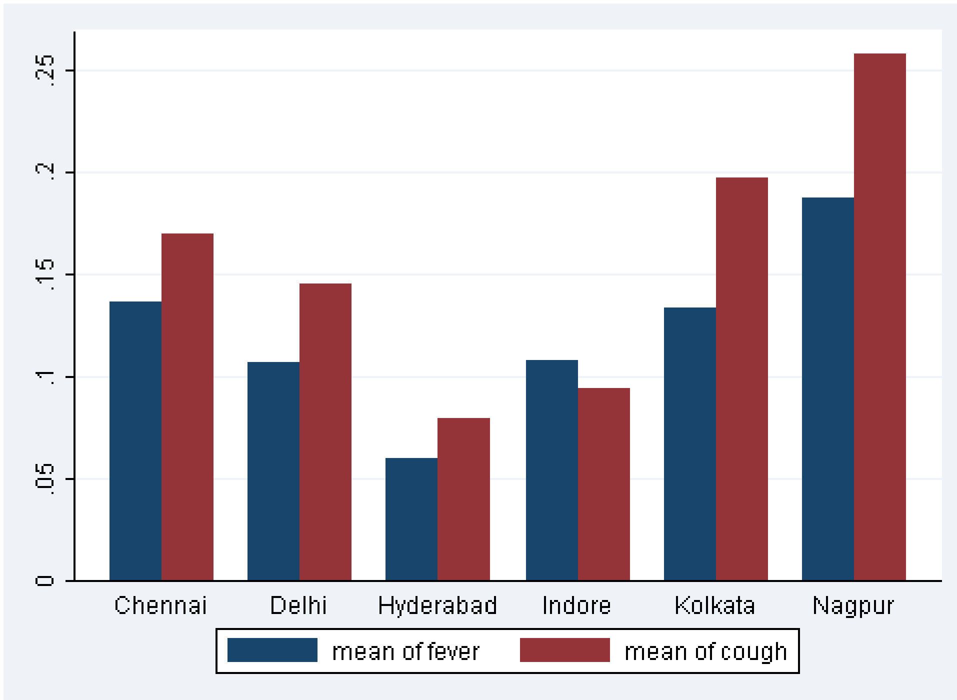
Source: Environmental Data Bank, MOEF, Govt. of India. Safe_SO2 is the level of SO2 that the MOEF advises as being safe. Each pollutant is measured in μm^3.

Figure 4 Monthly NO2 Distribution over NFHS 3 Interview Months across cities



Source: Environmental Data Bank, MOEF, Govt. of India. Safe_NO2 is the level of NO2 that the MOEF advises as being safe. Each pollutant is measured in μm^3.

Figure 2 City Level Prevalence of Child Health Outcomes



Source: NFHS 3 data

Figure 3 Probability (Fever, Cough) Vs Log (RSPM)

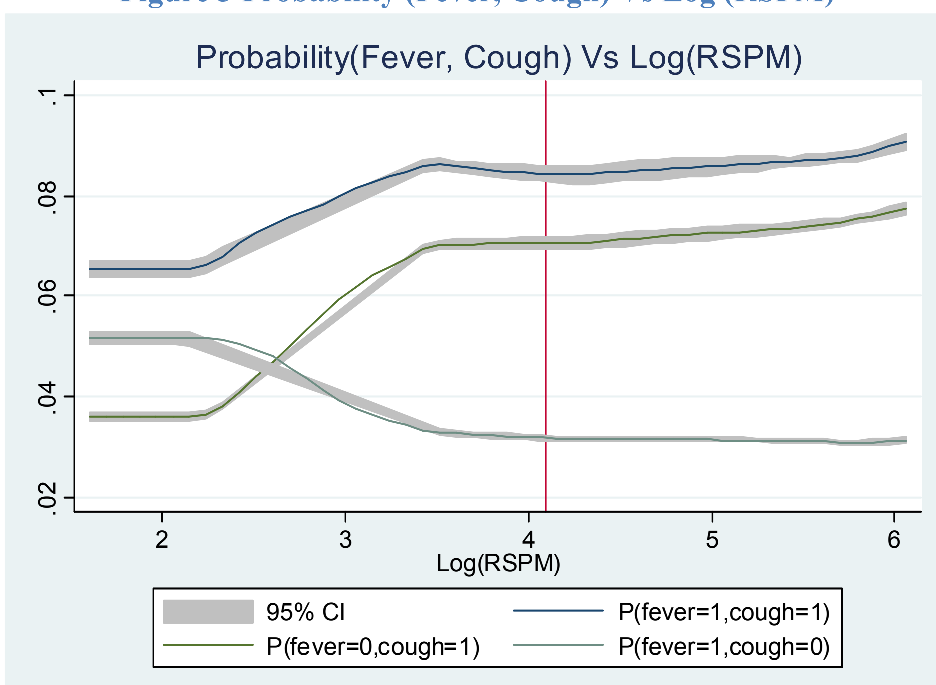


Figure 4 Probability (Fever, Cough) Vs RSPM Deviation from Safe Levels

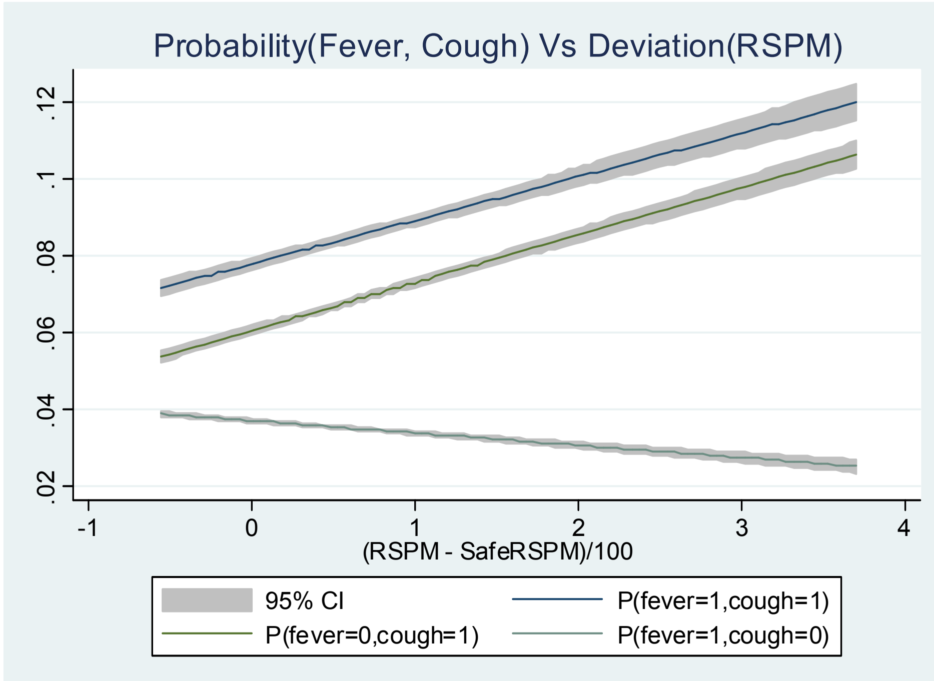
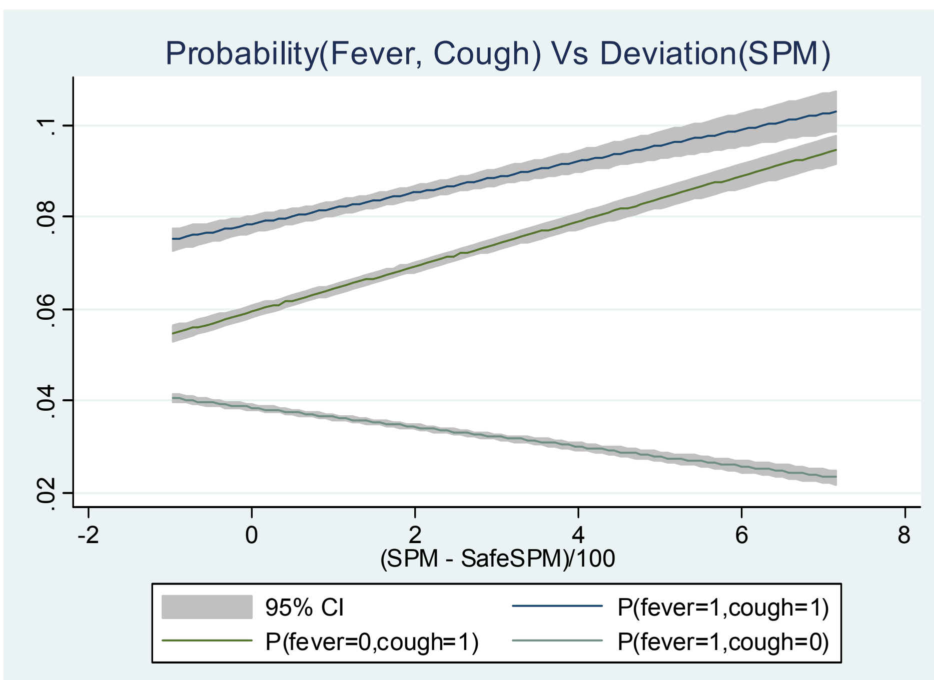


Figure 5 Probability (Fever, Cough) Vs Deviation in SPM from Safe Levels



Tables

Table 1: National Ambient Air Quality Standards (NAAQS)

S. No	Pollutant	Units	Time Weighted Avg.	Industrial, Residential and other Area	Ecologically Sensitive Area	Method of Measurement
1	Sulpher Dioxide (SO2)	μg/m^3	Annual	50	20	(1) improved west and Gaeke method;
			24 hours	80	80	(2) Ultraviolet Flurosence
2	Nitrogen Dioxide (NO2)	μg/m^3	Annual	40	30	(1) Modified Jacob & Hoechheiser (Na Arsenite)
			24 hours	80	80	(2) Chemiluminescence
2	Particulate Matter (PM10)	μg/m^3	Annual	60	60	(1) Gravimetric;
	(size < 10 μm)		24 hours	100	100	(2) TOEM; and (3) Beta attenuation
3	Particulate Matter (PM2.5)	μg/m^3	Annual	40	40	(1) Gravimetric;
	(size < 2.5 μm)		24 hours	60	60	(2) TOEM; and (3) Beta attenuation

Source: The Gazette of India, Nov 18th 2009. Available online at: http://www.cpcb.nic.in/National_Ambient_Air_Quality_Standards.php.

Table 2 Variance-Covariance Matrix for Alternative Measures of Outdoor Pollution

	SO21	NO21	RSPM1	SPM1
SO2	1			
NO2	0.7329	1		
RSPM	0.7167	0.9521	1	
SPM	0.7761	0.9276	0.9727	1

Table 3 Summary Statistics

Variable	N	Mean	SD	Min	Max
Uses Dirty Fuel (yes/no)	4657	0.220	0.414	0	1
Uses Unprocessed Fuel (yes/no)	4657	0.139	0.345	0	1
SO_2(µm)	4877	7.117	3.420	1.82	15.22
NO_2 (µm)	4877	34.849	23.558	4.61	102.05
RSPM (µm)	4877	114.454	97.190	4.96	429.96
SPM (µm)	4877	258.654	195.080	2.52	814.75
Child's age_(months)	4684	29.869	16.975	0	59
Child is Male? (yes/no)	4877	0.530	0.499	0	1
WHO Height for Age Z scores (HAZ)	4632	18.621	40.372	-6	99.99*
HAZ Flag	4877	0.238	0.426	0	1
WHO Weight for height Z scores (WHZ)	4632	19.028	40.164	-4.98	99.99*
WHZ Flag	4877	0.238	0.426	0	1
Asset Classes					
Poorest	4877	0.051	0.220	0	1
Middle	4877	0.156	0.363	0	1
Richer	4877	0.323	0.468	0	1
Richest	4877	0.470	0.532	0	1
Lives in a Slums	4864	0.407	0.491	0	1
Uses unsafe drinking water (yes/no)	4657	0.068	0.251	0	1
Uses unsafe toilet facilities (yes/no)	4651	0.070	0.255	0	1
Home has any windows? (yes/no)	4877	0.767	0.423	0	1
educlevel2	4877	0.116	0.320	0	1
educlevel3	4877	0.486	0.500	0	1
educlevel4	4877	0.178	0.383	0	1
SCST	4877	0.228	0.419	0	1
OBC	4877	0.295	0.456	0	1
Muslim	4877	0.219	0.414	0	1
Christian	4877	0.027	0.163	0	1
Other	4877	0.032	0.175	0	1

Note: Some of the HAZ and WHZ scores were implausibly high and for each such observation we have used a dummy variable to indicate that the value is too large for these variables. This helps keep about 24% of the sample for the analysis as opposed to dropping observations where the HAZ or WHZ were too large. Both these variables are quite tricky to measure and we believe that errors in this variable is unlikely to be informative about other variables and keep these observations and their flags that we also include in each of our specifications.

Table 4 RSPM and Indoor Air Pollution

VARIABLES	Model (1) Fever	Model (1) Cough	Model (2) Fever	Model (2) Cough	Model (3) Fever	Model (3) Cough	Model (4) Fever	Model (4) Cough
Indoor Proxies								
Dirty fuel (e.g. coal)	0.0271	0.147***	0.0152	0.107*	0.0118	0.0983*	0.0108	0.0990
	[0.0599]	[0.0557]	[0.0607]	[0.0563]	[0.0634]	[0.0597]	[0.0785]	[0.0738]
Unprocessed fuel (e.g. grass)	0.0567	0.148**	0.0501	0.119*	-0.0546	-0.00781	0.0278	0.0489
	[0.0710]	[0.0665]	[0.0710]	[0.0667]	[0.0719]	[0.0694]	[0.105]	[0.100]
Outdoor Measures								
Log(RSPM)			0.0322	0.104***	0.426***	0.581***	0.448***	0.609***
			[0.0207]	[0.0203]	[0.110]	[0.102]	[0.113]	[0.106]
Child Level								
Child Age (months)							-0.00581***	-0.00395***
							[0.00149]	[0.00141]
Child is Male? (0,1)							-0.0416	0.0157
							[0.0496]	[0.0476]
haz_who							-0.0304*	-0.00551
							[0.0177]	[0.0172]
whz_who							-0.0339	-0.0325
							[0.0212]	[0.0198]
Household Level								
Poorest							0.112	0.254
							[0.190]	[0.164]
Middle							0.131	0.0837
							[0.117]	[0.111]
Richer							0.0936	0.120*
							[0.0719]	[0.0695]
Lives in a Slum?							0.0335	0.0455
							[0.0573]	[0.0549]
Mother's Education								
educlevel2							0.184*	0.215**
							[0.0961]	[0.0907]
educlevel3							0.221***	0.281***
							[0.0806]	[0.0744]
educlevel4							0.120	0.225**
							[0.111]	[0.104]
Social Group								
SC or ST							-0.114	0.0334
							[0.0782]	[0.0703]
OBC							0.0988	0.0448
							[0.0688]	[0.0667]
Religion								
Muslim							0.0583	-0.0661
							[0.0696]	[0.0688]
Christian							0.142	0.0880
							[0.147]	[0.139]
Other							0.213	0.0772
							[0.135]	[0.133]
Month Counter	0.0301***	0.0154*	0.0351***	0.0280***	0.0206*	0.0119	0.0143	0.00228
	[0.00893]	[0.00884]	[0.00940]	[0.00894]	[0.0120]	[0.0113]	[0.0124]	[0.0118]
Observations	4463	4463	4463	4463	4463	4463	4415	4415
City Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes
$\hat{\rho}$= Cor(ε_1,ε_2)	0.835		0.837		0.828		0.831	

Note: Robust standard errors are in brackets below coefficient estimates; *** p<0.01, ** p<0.05, * p<0.1; omitted categories are "clean fuel", "richest", "no education", "general", and "Hindu", for indoor pollution, wealth quintile, mother's education, social groups, and religion respectively. The month counter is simply a count variable of which month the survey respondent was interviewed going from 1 to 11. Each regression has a constant, and dummies for levels of sanitation in the house (toilet, window, etc.). Finally, Wald Tests for $\hat{\rho} = 0$ are rejected at p values < 0.001.

Table 5 Effect of Other Ambient Air Pollution Measures

	Model (5)		Model (6)		Model (7)	
VARIABLES	Fever	Cough	Fever	Cough	Fever	cough
Indoor Proxies						
Dirty fuel (e.g. coal)	0.0217	0.109	0.0155	0.103	0.0115	0.100
	[0.0785]	[0.0739]	[0.0784]	[0.0739]	[0.0784]	[0.0738]
Unprocessed fuel (e.g. grass)	0.0349	0.0564	0.0316	0.0530	0.0305	0.0523
	[0.105]	[0.100]	[0.105]	[0.0999]	[0.105]	[0.100]
Outdoor Measures						
Log(SO2)	0.630***	0.631***				
	[0.208]	[0.187]				
Log(NO2)			0.0619	0.0958		
			[0.124]	[0.112]		
Log(SPM)					0.516***	0.915***
					[0.177]	[0.167]
Child Level						
Child Age (months)	-0.00576***	-0.00388***	-0.00577***	-0.00387***	-0.00577***	-0.00388***
	[0.00149]	[0.00141]	[0.00148]	[0.00141]	[0.00149]	[0.00141]
Child is Male? (0,1)	-0.0438	0.0126	-0.0431	0.0133	-0.0414	0.0168
	[0.0495]	[0.0473]	[0.0494]	[0.0473]	[0.0495]	[0.0475]
haz_who	-0.0312*	-0.00734	-0.0321*	-0.00828	-0.0310*	-0.00600
	[0.0177]	[0.0171]	[0.0177]	[0.0170]	[0.0178]	[0.0172]
whz_who	-0.0325	-0.0294	-0.0314	-0.0286	-0.0329	-0.0326*
	[0.0213]	[0.0196]	[0.0212]	[0.0195]	[0.0212]	[0.0197]
Household Level						
Poorest Asset Quartile	0.128	0.270*	0.123	0.268*	0.114	0.258
	[0.189]	[0.163]	[0.189]	[0.162]	[0.190]	[0.164]
Middle Asset Quartile	0.139	0.0917	0.135	0.0899	0.134	0.0925
	[0.117]	[0.110]	[0.116]	[0.110]	[0.117]	[0.111]
Richer Asset Quartile	0.0986	0.127*	0.0979	0.126*	0.0945	0.120*
	[0.0719]	[0.0695]	[0.0719]	[0.0693]	[0.0718]	[0.0695]
Household lives in a Slum?	0.0126	0.0178	0.0139	0.0208	0.0283	0.0414
	[0.0567]	[0.0542]	[0.0573]	[0.0548]	[0.0573]	[0.0549]
Mother's Education						
educlevel2	0.177*	0.205**	0.176*	0.205**	0.176*	0.202**
	[0.0957]	[0.0901]	[0.0954]	[0.0900]	[0.0958]	[0.0909]
educlevel3	0.218***	0.273***	0.214***	0.271***	0.215***	0.273***
	[0.0800]	[0.0737]	[0.0799]	[0.0738]	[0.0803]	[0.0749]
educlevel4	0.110	0.208**	0.106	0.206**	0.119	0.230**
	[0.110]	[0.103]	[0.110]	[0.103]	[0.111]	[0.104]
Social Group						
SC or ST	-0.103	0.0441	-0.108	0.0415	-0.112	0.0378
	[0.0781]	[0.0700]	[0.0781]	[0.0700]	[0.0782]	[0.0703]
OBC	0.0984	0.0438	0.100	0.0461	0.0941	0.0376
	[0.0689]	[0.0670]	[0.0687]	[0.0669]	[0.0690]	[0.0669]
Religion						
Muslim	0.0773	-0.0399	0.0770	-0.0387	0.0610	-0.0652
	[0.0689]	[0.0680]	[0.0688]	[0.0680]	[0.0693]	[0.0687]
Christian	0.153	0.102	0.154	0.109	0.156	0.113
	[0.148]	[0.139]	[0.147]	[0.138]	[0.147]	[0.140]
Other	0.217	0.0817	0.217	0.0834	0.212	0.0761
	[0.134]	[0.133]	[0.134]	[0.133]	[0.134]	[0.133]
Month Counter	0.0404***	0.0367***	0.0349***	0.0299***	0.0183	6.89e-05
	[0.0107]	[0.0102]	[0.0118]	[0.0112]	[0.0127]	[0.0120]
Observations	4415	4415	4415	4415	4415	4415
$\hat{\rho}$= Cor(ε_1,ε_2)	0.832		0.833		0.832	

Note: See Footnote for Table 1. In addition, all models have city level fixed effects.

Table 6 Effect of Deviating from Mandated Safe Levels of Ambient RSPM and SPM

	Model (8)		Model (9)	
VARIABLES	Fever	Cough	Fever	Cough
Indoor Proxies				
Dirty fuel (e.g. coal)	0.0130	0.0997	0.00983	0.0973
	[0.0784]	[0.0740]	[0.0783]	[0.0737]
Unprocessed fuel (e.g. grass)	0.0267	0.0464	0.0352	0.0596
	[0.104]	[0.100]	[0.105]	[0.0999]
Outdoor Measures				
dRSPM = (RSPM-μ_{RSPM})/100†	0.162***	0.235***		
	[0.057]	[0.053]		
dSPM = (SPM-μ_{SPM})/100†			0.104***	0.154***
			[0.0387]	[0.0364]
Child Level				
Child Age (months)	-0.00577***	-0.00387***	-0.00576***	-0.00386***
	[0.00148]	[0.00141]	[0.00149]	[0.00141]
Child is Male? (0,1)	-0.0452	0.0107	-0.0414	0.0157
	[0.0494]	[0.0473]	[0.0495]	[0.0475]
haz_who	-0.0319*	-0.00779	-0.0318*	-0.00765
	[0.0177]	[0.0170]	[0.0178]	[0.0171]
whz_who	-0.0322	-0.0293	-0.0333	-0.0317
	[0.0212]	[0.0195]	[0.0212]	[0.0196]
Household Level				
Poorest Asset Quartile	0.121	0.265	0.124	0.272*
	[0.189]	[0.162]	[0.190]	[0.163]
Middle Asset Quartile	0.131	0.0847	0.132	0.0866
	[0.116]	[0.110]	[0.117]	[0.111]
Richer Asset Quartile	0.0978	0.127*	0.0949	0.121*
	[0.0720]	[0.0695]	[0.0718]	[0.0693]
Household lives in a Slum?	0.0120	0.0174	0.0295	0.0424
	[0.0568]	[0.0542]	[0.0575]	[0.0551]
Mother's Education				
educlevel2	0.176*	0.206**	0.177*	0.206**
	[0.0953]	[0.0900]	[0.0958]	[0.0906]
educlevel3	0.212***	0.270***	0.219***	0.280***
	[0.0797]	[0.0738]	[0.0804]	[0.0744]
educlevel4	0.101	0.200*	0.119	0.226**
	[0.110]	[0.104]	[0.110]	[0.104]
Social Group				
SCST	-0.109	0.0386	-0.106	0.0444
	[0.0781]	[0.0701]	[0.0782]	[0.0703]
OBC	0.105	0.0523	0.100	0.0468
	[0.0687]	[0.0669]	[0.0688]	[0.0669]
Religion				
Muslim	0.0779	-0.0378	0.0710	-0.0470
	[0.0688]	[0.0680]	[0.0691]	[0.0684]
Christian	0.149	0.0982	0.156	0.109
	[0.147]	[0.138]	[0.147]	[0.139]
Other	0.218	0.0840	0.218	0.0864
	[0.134]	[0.133]	[0.134]	[0.133]
Month Counter	0.0359***	0.0317***	0.0161	0.00234
	[0.0108]	[0.0102]	[0.0138]	[0.0132]
Observations	4415	4415	4415	4415
$\hat{\rho}$= Cor(ε_1,ε_2)	0.832		0.832	

Note: † μ_{RSPM} & μ_{RSPM} indicate d NAAQS defined save level for RSPM and SPM as mentioned in Table 1. dRSPM and dSPM are scaled deviations from these levels to capture the health impacts of violating the safety norms. Read footnote for Table 2 for further clarifications.

Table 7 Predicted Probabilities at the Mean and Maximum of the Outdoor Air Pollution Distribution

	Log (RSPM)		Dev (RSPM)		Dev (SPM)	
	Mean	Max	Mean	Max	Mean	Max
Pr(Fever = 1, Cough = 1)	0.119	0.324	0.063	0.152	0.088	0.240
Pr(Fever = 1, Cough = 0)	0.030	0.027	0.033	0.037	0.035	0.030
Pr(Fever = 0, Cough = 1)	0.069	0.164	0.050	0.113	0.068	0.173
Pr(Fever = 0, Cough = 0)	0.782	0.485	0.854	0.697	0.809	0.557