Research Article

Exposure to HAP and the Regional Pattern of Air-Related Morbidity in India: A Multivariate Analysis

Sabitri Dutta

DumDum Motijheel Rabindra Mahavidyalaya, 208/B/2 DumDum Road, Kolkata 700074, India

Correspondence should be addressed to Sabitri Dutta; sabitri.d@gmail.com

Received 26 August 2014; Revised 4 June 2015; Accepted 23 June 2015

Academic Editor: Alberto Davila

Copyright © 2015 Sabitri Dutta. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper attempts to analyse the nature of relationship between the levels of Household Air Pollution (HAP) and incidence of morbidity in urban India along with the pattern of its regional variations. It also explores the causal connection between the incidence of air-related diseases and HAP, giving control for outdoor pollution. Whether these effects are sensitive to the overall level of economic development is an issue of our interest. Therefore, the States/UTs of India are grouped, in terms of different developmental parameters, using Multidimensional Scaling and Clustering technique, and the relation between HAP and morbidity has been analysed for each group of states by applying stepwise regression techniques. The study of this heterogeneity helped us provide more focused policy prescriptions. Different policy prescriptions in terms of education as well as exposure to HAP and fuel choices have been suggested for different clusters.

1. Introduction

Household Air Pollution (HAP) has a serious health outcome, causing annually 3.3% deaths and 2.7% morbidity worldwide [1]. In the high mortality developing countries (which make up 40% of the world population), HAP is considered to be the 4th most dangerous killer after malnutrition, unsafe sex, lack of safe water, and sanitation. This paper tries to identify the possible causes that may explain the indoor air-related health hazards in the urban parts of the Indian subcontinent. In the urban settings, Outdoor Air Pollution, or OAP, can be an important contributor to indoor air quality [2, 3]. Exposure to OAP largely depends on the type of economic activities the victim engages in. The economic engagements are directly related to the level of development. Moreover, exposure to HAP primarily depends on the fuel types used in the household, the mode of cooking, living condition (structure of the house, ventilation, etc.), time spent near the pollution source, vulnerability of the victims, and so forth. These factors are reflected by the income standard, educational attainment, and health status of the person exposed. They have direct relation with the developmental achievements of the exposed individuals. In fact, these are the development indicators

that are used in the measurement of Human Development Index. Therefore, the problem of HAP, as a function of OAP, or the problem itself, has an obvious connection with economic development and needs to be addressed from solely an economic perspective. So Indian States/UTs, with homogeneous developmental achievements, in terms of the factors mentioned above, are clubbed together for more focused and effective policy suggestions. This grouping of regions is also important, as the existing literature supports a huge disparity in the exposure to HAP between the developed and developing parts of the world. The problem of particulate matter pollution is more severe in the developing countries, accounting for 90% of the global exposure, leaving only 10% for their developed counterpart [4, 5].

Indian urbanization became more active from the beginning of 21st century. The share of the Indian population living in urban areas increased from around 28% (290 million) in 2000 to around 30% (340 million) in 2008 and is expected to increase to 40% (590 million) by 2030 [6]. Most of these Indian cities are characterized by large slum-dwelling population, according to the National Family Health Survey-3 [7]. Poverty is more prevalent in the slum areas than their nonslum counterparts. However, the NFHS report says that

the proportions of the poor living in nonslum areas are substantial in virtually all the major cities of the country. The occupational structure of both women and men is quite diversified in these cities. Among the urban males in India, 21% are employed in the formal sector and 79% are employed in the informal sector as of 2004-05. For urban females, formal employment was 16%, and informal part was as large as 84%. In general, women workers in the slum areas of every city are concentrated more in the production and service activities, whereas, women workers in the nonslum areas work more in the production and professional activities. Poor women workers are mostly engaged in service-related and production activities. Slums have much poorer housing conditions when compared to nonslum areas in terms of construction material, residential crowding, or ventilation of the dwelling. However, the poor have the worst housing conditions in all counts. The accessibility to improved toilet facilities is not very high in most of these cities. In almost all the cities, the accessibility to proper sanitation facility is much worse in the slum areas when compared to those in the nonslum areas. Open defecation is the highest among the poor in every city in India. Literature says that the condition of rural and urban poor is the same in terms of exposure to HAP [8] as the latter cannot always afford clean fuels, which are less polluting in nature. With this background, a study has been carried out only for the urban Indians as the exposure to HAP for urban fringes has become a neglected part of the research, where a limited amount of research is available, which analyses the problems of the rural areas.

2. Aims and Objective

In this paper, three different exercises at different levels would be reported. In the first exercise, the relation between general morbidity and HAP is analysed to construct an idea about the influence of indoor pollution on the overall morbidity level. Then, the relation of HAP with prevalence of air-related morbidity, specifically, is verified. Locating the disparity among the Indian states, in terms of health impact of HAP, is the second level of exercise. Effective policy suggestions are subsequently discussed in the third exercise.

The rest of the paper is organized as follows: the following section describes the National Sample Survey Organisation (NSSO) data used for the study to form the quantitative background of the analysis. The subsequent section describes the methodology and tries to analyse the relationship between the prevalence of general and air-related morbidity and the type of fuel used for controlling a number of socioeconomic characteristics at the individual level for urban India as a whole. Relevant methodology for locating the disparities among Indian states and Union Territories in terms of developmental indicators is explained in this section. Results and discussions are explained in the next section. Exploratory data analysis has been carried out here to verify the statistical properties of all relevant variables across different groups of States/UTs and to decide upon the transformations that are needed to make them amenable to regression analyses. The section reports the regression results to isolate the marginal influence of the exposure level and the extent of morbidity

by providing control to all other associated factors at the individual level. The final section concludes the paper with some state-specific policy suggestions for combating HAP.

3. Data

The 60th round household survey data on morbidity and health care (Schedule 25.0) by NSSO provides the information required for the analysis. The relevant information from NSSO that has been used in the study includes the monthly per capita expenditure (that is used as a proxy for income as no direct data is available), the energy used for cooking, and the education level of the individual as well as the information on morbidity at the individual level: whether the individual is ailing over the last 15 days before the survey (yes/no type) or not. (It has been established by NSSO through several experiments carried out on optimal recall period that, for temporary indisposition, the recall is most reliable over the last 15 days only. The remote memory of any temporary ailment is not retained over a longer period, barring the exception of a hospitalization episode. So, here, for the hospitalization episode, the reference period has been taken as 1 year and, for other ailments, the last 15 days.) In the first level of exercise, the relation between the general morbidity and the indoor pollution has been verified. For that, the relevant data are made available on 131,369 individuals for the Indian urban sector as a whole. The fuel-based energy that is consumed in the households has been used as a proxy for HAP as no direct data on indoor concentrations of pollutants are available. In the next stage, as our purpose is to analyse the health-related effects of indoor pollution, only the airborne diseases, available in the data source, have been considered. These include respiratory disease, tuberculosis, asthma, neurological disorders, conjunctivitis, cataract, skin diseases and anaemia [9–12]. The ailments in this case are self-reported morbidity made by the informant. (It may be noted that some of the ailments may be treated (either as an inpatient of a hospital or otherwise) and some will remain untreated; both cases are considered here. A person under medication for an ailment during the reference period, whether he/she felt sick or not, is treated as ailing, and cases of complications arising during pregnancy or after childbirth are considered to be ailments. However, untreated injuries, like cuts, burns, scald, and bruise of minor nature are not considered to be ailments unless the informant considers them to be severe enough.) For this case, 9,556 urban individuals have been retained on whom all the relevant information were available. All these individuals have reported themselves to be morbid (in air-related or other type diseases) in the last 15 days before the survey was conducted. This information is then spread over 28 states and 7 Union Territories. Instead of fuel type, the emission factors for each of the fuels used have been used in this case to account for the pollution potential of the fuels available in the data source. The emission factors have been taken directly from Smith et al. [9] (Figure 6).

Moreover, the vulnerability to the risk of being affected by HAP-induced diseases is likely to vary with the age and sex of the victim involved. In order to consider this, vulnerability weights, for individuals of different age and sex, are used.

TABLE 1: Age-sex based relative vulnerability weights.

Adult male	Adult female	Old male	Old female	Children
1.0	1.0	1.8	1.2	1.6

Source: Kathuria and Khan [13].

These weights have been taken from Kathuria and Khan [13], where it was proposed that a 5-hour exposure should be provided for adult males and adult females, 6 hours should be provided for old females, 8 hours should be provided for children, and 9 hours should be provided for old males. Considering the exposure rate of adult males and females to be the norm, we have calculated the vulnerability weights for each age-sex based group, as shown in Table 1.

4. Methodology

4.1. Individual Level Morbidity and Socioeconomic Factors: All India (Urban) Analysis

4.1.1. Variable Construction and Hypothesis. At first, attempts were made to explain the general morbidity of the individuals in terms of a number of socioeconomic variables including the pattern of energy use (proxy for HAP). The variables are defined as follows:

M. Incidence of general morbidity at the individual level (ailing over last 15 days; yes/no type, taken directly from the data source): this is the study variable in this level of analysis.

MPCE. Monthly per capita expenditure of the household: this is directly taken from the database. It is expected to have a negative effect on the M.

FUEL. A dummy variable has been defined with clean fuel (Electricity, LPG, and Kerosene) represented by 0 and dirty fuel (Firewood, Biomass, and Coal) represented as 1. So, the variable FUEL, as defined, is expected to have a positive effect on the level of morbidity.

VULN. Age and sex-based weights as shown in Table 1: this should also have a positive impact on the morbidity.

EAI. It stands for the level of education of the individual. To simplify the matter, we have defined an educational attainment index for each individual as $EAI = (\text{Actual}/\text{Maximum}) \times 100$.

For children 0–6 years of age, the education level of the head of the respective households has been used as the children of this age group do not have access to education on their own. This variable is expected to have a negative effect on the study variable.

In the next step, attempts are made to find an association specifically between air-related morbidity and HAP. HAP-induced morbidity depends primarily on the level of exposure and also on socioeconomic influences, like income and awareness. We define the following variables for this analysis:

ABD. Individual level incidence of air-related diseases, where if there is any air-related morbidity episode reported by the individual over the reference period (15 days before survey),

then ABD is assigned a value 1 and 0 otherwise. Thus, ABD is a dummy variable. This is the dependent variable in this exercise.

MPCE. Same as before.

EMSN. Emission factors from different types of fuels (Figure 6); it is expected to have a positive impact on ABD.

VULN. Same as before.

EXP. The exposure for each individual based on emission and vulnerability weights. This has been calculated as

$$EXP_i = (\text{emission factor of the fuel used by the household that the individual belongs to}) * (\text{the age-sex base vulnerability of that particular individual}) = (EMSN_i * VULN_i);$$

EXP is expected to have a positive influence on air-related morbidity.

EAI. Same as before.

The main method of analysis will be through exploration, correlation analysis, stepwise regression analysis, Multidimensional Scaling Technique, and Cluster Analysis.

Here, a small account is presented about the position of the urban dwellers in terms of some of their morbidity patterns and some socioeconomic variables that have been mentioned above. The NSSO data that were used in the analysis came up with some important findings. For the urban people of India, morbidity is the highest for the high MPCE class. Possibly, this is due to the fact that this morbidity is entirely the respondent's perception, and people belonging to the higher MPCE class are likely to have better awareness regarding the diseases they suffer from and have greater affordability of getting treatment. Hence, they have more reported morbidity. The economically weaker section of the society does not always recognize this discomfort to be a disease and hence does not report it as an episode of morbidity. The greater prevalence at the higher income class is observed for air-related diseases too. The reason remains the same as before. As far as general living condition is concerned, the NSSO provides information on housing structure, drainage type, latrine type, and so forth at the household level. In the urban areas, most of the structures are *pucca* (concrete structure) type, and, therefore, the concentration of morbid people is higher in the households with a *pucca* structure. However, as per the received wisdom, underground drainage and septic/flush system should ensure better environmental quality, leading to lesser pollution-induced illness; contrary to our expectation the morbidity is surprisingly high for the individuals in the households with underground drainage systems and with pit latrines followed by the septic tank/flush system. The morbidity prevalence is the highest among those individuals whose source of drinking water is a *pucca* (concrete) well. The extent of morbidity, especially when it is air-related, is expected to be higher in case of dirty fuels. However, data reveal a different picture, where the air-related disease is more in the households, using relatively cleaner fuels.

These findings make us feel uncomfortable, especially regarding the quality of NSS data and/or the complexity of the underlying influences, which affect the apparently simple relation between pollution exposure and the incidence of morbidity. So, it would be more pertinent to explore the inherent relationship of the pattern of morbidity, the type of fuel used, and also the exposure (Since the analysis has been carried out only on the basis of fuel usage of the household, as an indicator of indoor air pollution, the seasonality factor is not taken care of. Moreover, the calculation of exposure to HAP involves the emission factors of the fuels that are used by the household. Here, seasonality is not of much relevance. However, the outdoor air quality does have an influence on the indoor environment as mentioned by various literatures [14]) to HAP by controlling all the other related influences, like economic standing and level of education.

It should be mentioned here that morbidity of any kind is largely dependent on the living condition of the household where the individual stays. Ideally, this should include the ventilation inside the house, the size of the room, congestion, and many other factors. Moreover, apart from the type of fuel used, indoor pollution also depends on the kind of stoves that are used for cooking and other purposes, the distance of the ventilation source from the cooking hearth, the location of the kitchen inside the house, and so forth. Information about these factors is not available in the NSSO data source. So, the present study is unable to include the living conditions of the people as an explanatory variable for HAP-induced diseases. As income is a factor which may ensure the affordability for better living condition, it is thought to explain some of the variations in the incidence of M or of ABD due to the change in the living conditions.

5. Result and Discussion

5.1. Data Exploration. Prior to detailed analysis, an exploration of the data has been carried out here to get an idea about the nature of the data structure. In the first step, the sample distributions of all the variables (except the binary variables: M, FUEL, and ABD) are scrutinized in terms of their descriptive statistics (Table 2).

Since for linear regression to yield good estimates we need all variables to conform to the normal distribution and hence depending on the descriptive measures, we check whether the data is reasonably symmetric and, if so, whether the tails are sufficiently thin or not. The variables are then transformed accordingly. Only MPCE undergoes a square root transformation, and EXP takes an inverse square ($1/\text{square}$) transformation. The transformed variables would be denoted as $\sqrt{\text{MPCE}}$ and $1/\text{sqEXP}$. The details of the exploration procedure and the basis of transformation are discussed in the appendix.

5.2. Analysis of Correlation and Check for Multicollinearity. The pairwise correlation between all the selected and transformed variables has been analysed at the next stage (Table 3). For general morbidity, its relation with vulnerability and educational attainment is coming as expected. Confusion

arises when the sign of association of morbidity with $\sqrt{\text{MPCE}}$ and FUEL contradicts the expectation. The statistically significant positive linear association between $\sqrt{\text{MPCE}}$ and M is counterintuitive. Here, two plausible explanations may be provided: (a) with the increase in the affordability via the higher level of income, the perception may change and reported illness of the household may go up; and (b) the relationship between the reduction in morbidity and a higher income level may be effective after some threshold level of achievement. Fuel type may not be an appropriate proxy for HAP. Again, HAP is not likely to explain all types of morbidity in general. This may be the possible reason why the expected sign is not obtained for association between the fuel type and the general morbidity level.

As no significant correlation is found between the fuel type and general morbidity, the correlation is checked for the air-related diseases and HAP. Instead of fuel type, here the emissions from different fuels and also the exposure to HAP have been taken into consideration. Statistically strong association with expected signs is found for the incidence of air-related diseases (ABD) with all the explanatory variables. The sign of association of EMSN with $\sqrt{\text{MPCE}}$ and EAI conforms our expectation that with the increase in income and improvement in the education level fuel emission reduces, implying a switch from dirty to clean fuel. Contrary to our expectation, as income improves, vulnerability increases. However, the association of exposure with both of these variables is found as expected with significance.

Since the pairwise correlation between all explanatory variables is statistically significant, the possibility of multicollinearity needs to be verified. The Variance Inflation Factor (VIF), tolerance level ($1/\text{VIF}$), and the Condition Number (CN) are checked for all the variables. This is shown in Table 4 (for only air-related morbidity case) (for general morbidity case, it was checked, and no problem could be recognized).

As a rule of thumb, a variable whose VIF values are greater than 10 and overall CN higher than 30 may merit further investigation. Since the VIF factors are lying between 1.01 and 3.23, with a mean value of 1.82, that is, all the VIF values are much below 10 and the CN is below 30 (20.10), therefore the threat of multicollinearity is not very compelling, and the standard regression exercise can be carried out on the data set. It should be mentioned here that the EXP is a function of EMSN and VULN by definition and VIF is rated highest here for the variable with the lowest tolerance. Obviously, in the subsequent analysis, EXP would not be taken while EMSN and VULN are used and vice versa.

5.3. Regression Analysis. Two sets of regressions are run with two representations of morbidity, which are general and air-related (M and ABD). Since the dependent variables are dummy variables of binary nature (taking only values “0” and “1”), the *PROBIT* method of regression technique has been considered to be appropriate. In order to identify the most appropriate variables to enter into the model serially, we choose a stepwise regression procedure, where the model endogenously selects the variables to be added in each step. The level of significance for a variable to be added in the

TABLE 2: Descriptive statistics (all India).

(a) General morbidity			
Descriptive statistics	MPCE	VULN	EAI
Mean	917.2	1.217	56.64
Median	757.1	1.00	66.67
Standard deviation	685.9	0.29	34.35
Pseudo SD [#]	416.2	0.44	37.06
Sample size	132369	132369	132369
Required transformation	Square root	Takes certain values, no transformation	No transformation

[#] A pseudo standard deviation is IQR/1.35 that should be equal to the actual standard deviation for normal density.

(b) Air-related morbidity					
Descriptive statistics	MPCE	EMSN	VULN	EXP	EAI
Mean	1037.22	8.87	6.39	56.30	49.93
Median	833.33	1.00	5.00	8.00	50
Standard deviation	833.73	14.89	1.65	98.71	36.75
Pseudo-SD	503	18.53	2.22	92.66	61.77
Sample size	9556	9556	9556	9556	9556
Required transformation	Square root	Takes certain values, no transformation	Takes certain values, no transformation	1/square	No transformation

Source: calculated from the NSS dataset, 60th Round.

TABLE 3: Pairwise correlation.

(a) General morbidity					
Variables	M	sqrtMPCE	FUEL	VULN	EAI
M	1.0000				
sqrtMPCE	0.0637*	1.000			
FUEL	-0.0062*	-0.4136*	1.000		
VULN	0.0642*	-0.0787*	0.0491*	1.000	
EAI	-0.0628*	0.3881*	-0.3438*	-0.1814*	1.000

(b) Air-related morbidity						
Variables	ABD	EMSN	VULN	1/sqEXP	sqrtMPCE	EAI
ABD	1.0000					
EMSN	0.0517*	1.000				
VULN	0.0217*	-0.0655*	1.000			
1/sqEXP	-0.0504*	-0.6406*	-0.1427*	1.000		
sqrtMPCE	-0.0439*	-0.3103*	0.1561*	0.3415*	1.000	
EAI	-0.0595*	-0.2862*	-0.0842*	0.3141*	0.4338*	1.000

Source: calculated from the NSS dataset, 60th Round.

*Significant at 5% level.

model has been taken as 10%. The result is shown in Table 5. From this table, a few observations can be made. The value of LR $\chi^2(k)$ is significant at less than 1% level for all cases, making all regressions statistically acceptable. Regressions (1) and (2) have taken the incidence of morbidity (M) as the study variable, where it is air-related morbidity (ABD) in both regressions (3) and (4).

For each type of morbidity, two regressions have been run and reported as R_1 and R_2 for M and R_3 and R_4 for ABD. Regression R_1 endogenously includes VULN, sqrtMPCE,

and EAI stepwise and leaves out FUEL, which is the main variable of interest. This regression shows that the general morbidity increases as the victim becomes more vulnerable in terms of age and sex, and this morbidity depends upon the awareness of the individual captured in terms of his/her educational attainment. However, some discomfort creeps in as the sign of sqrtMPCE is positive and significant. This may be due to the presence of some degree of nonlinearity in the actual relationship. That means the effect of improvement in income will be reflected in lowered morbidity provided that

TABLE 4: Variance inflation factor and tolerance.

Variable	VIF	SQRT VIF	Tolerance	R-squared
ABD	1.01	1.00	0.99	0.005
EMSN	2.23	1.49	0.45	0.551
VULN	1.81	1.35	0.55	0.447
1/sqEXP	3.23	1.80	0.31	0.690
sqrtMPCE	1.34	1.16	0.74	0.255
EAI	1.31	1.14	0.76	0.237
Mean VIF: 1.82		Condition number: 20.0990		

the improvement is substantive. This means that there may exist some threshold levels of MPCE after which they would be causally effective to reduce M. This was also reflected in the correlation results obtained in Table 3.

To verify this possibility, we have additionally introduced MPCE as another regressor. The results are reported under Regression 2. For MPCE, the variable has significant influence on the incidence of morbidity both at the square root and level with opposite but expected signs. The effect of the increase in income is achieved not at the very initial stage but after a threshold level. As MPCE increases, M increases up to a certain level, but, after a substantive increment, the effect gets reversed. The relationship is inverted “U” shaped. After a certain (threshold) level of income, an improvement in the expenditure leads to a better health status. At this stage, the significant effect of FUEL is obtained with expected signs confirming its influence on the general health of the public at the individual level.

In the next set of regressions, the air-induced morbidity is taken as the study variable and to be more specific instead of fuel emissions from different fuels and the resulting exposure of the victims are considered. The results are reported in R.3 and R.4. In R.3, EAI is included first and explains ABD with less than 1% significance with the expected sign. In the second and third steps, EMSN and VULN are added, respectively. Expected signs, with statistical significance, come for both variables. Confusion arises when sqrtMPCE fails to be included as an explanation for ABD, contradicting the hypothesis made on the variable. In the following regression (R.4), instead of EMSN and VULN, the exposure variable (1/sqEXP) is considered, combining both emission and vulnerability factors. Here, in the first step, EAI is added, and an expected relation is obtained with strong statistical significance. In the second step, 1/sqEXP is included, and a statistically strong inverse relation is found with the incidence of ABD which actually implies a direct relationship between the incidence of air-related disease and the exposure to HAP. Like the previous case, again, the model does not include sqrtMPCE. In both cases, HAP, either as emission or as exposure, confirms that it has a substantial impact on the incidence of air-related morbidity in the urban part of India.

The fact that the income variable fails to explain the air-induced morbidity raises discomfort, although it has a nonlinear impact on general morbidity. But it is unlikely that air-related morbidity does not get affected as the income level

of the individual improves. The health status generally gets better off with the advancement of developmental parameters, where income is the most important factor. But the absence of such relation for the air-related illness implies that there is some omitted variable problem which captures the income factor. This happens possibly because air-related morbidity is a combined effect of both indoor pollution and outdoor pollution, and it is not explained by pollution exposure from a single source. Income of an individual depends largely on his/her outdoor exposure, and, as the role of the ambient pollution is omitted here, the effect of income fails to be obtained.

Apart from the factors like emission, vulnerability, or exposure, which are directly related to this type of morbidity, the sign and strong association of ABD and EAI implies a strategic need to spread awareness among the people about the extent of the harm and hazard that may be caused by in-house pollution. Exposure and education level are two areas where improvements can be brought forward. Exposure, in the present analysis, is defined as a combination of vulnerability and emission, of which the former again depends on the age and sex of the exposed individual and is not subject to policy parameter. Emission, on the other hand, depends on the type of fuel used. So, the only factor that can reduce exposure is the use of safer fuel and better stoves to improve the emission factor. Now, the change in fuel type is a household decision, and it is very difficult to change the household factors as it involves the household's taste and preferences, culture, belief, and certain age-old habits. If the change in fuel type is targeted through a proper subsidization scheme, the effect may not be as satisfactory as it is expected [15].

Moreover, even if it is possible to change the fuel at the household level, the efficiency of it depends upon the type of cooking hearth that the household uses. For that, the Improved Cook Stove (ICS) Program has been considered to be an important policy strategy, suggested by many literatures. Though the RESPIRE study in Guatemala has worked well, the success of the ICS Program was limited because of the various social and physical factors that hindered the use of these stoves altogether. Behavioral factors became all important in the context of such interventions.

Any behavioral change depended upon the awareness of the individuals especially of the victims and of the main decision-maker in the household. This awareness could be improved through proper educational training right from the primary level of schooling. So, at the other end, effective policy design includes improvements made in attainment of education. Given the federal structure of India, any intervention in the areas like health, education, and environment is a state subject. But India is a country with considerable geographical, social, cultural, and economic disparities. The factors where different Indian states differ include, among others, income, education, living standard, health, and infrastructural facilities. So, to plan any intervention, state-specific factors should be taken into consideration. Depending on the issue under consideration, the states may be grouped under different clusters endogenously in accordance with their intrinsic similarities or their developmental attainments.

TABLE 5: Regression results for HAP and morbidity: all urban India.

	Dependent variable M		Dependent variable ABD			
Regressors and diagnostic values	$p = 0.0000 < 0.1000$ adding VULN	$p = 0.0000 < 0.1000$ adding VULN	$p = 0.0000 < 0.1000$ adding EAI	$p = 0.0000 < 0.1000$ adding EAI	$p = 0.0000 < 0.1000$ adding EAI	$p = 0.0000 < 0.1000$ adding EAI
	$p = 0.0000 < 0.1000$ adding sqrtMPCE	$p = 0.0000 < 0.1000$ adding sqrtMPCE	$p = 0.0000 < 0.1000$ adding EAI	$p = 0.0006 < 0.1000$ adding EMSN	$p = 0.0012 < 0.1000$ adding 1/sqEXP	$p = 0.0012 < 0.1000$ adding 1/sqEXP
	$p = 0.0000 < 0.1000$ adding EAI	$p = 0.0000 < 0.1000$ adding MPCE	$p = 0.0172 < 0.1000$ adding VULN			
	$p = 0.0000 < 0.1000$ adding FUEL	$p = 0.0306 < 0.1000$ adding FUEL				
	R.1	R.2		R.3		R.4
VULN	0.3185***	0.3211***	EAI	-0.0020***		-0.0019***
sqrtMPCE	0.0191***	0.0306***	EMSN	0.0035***		—
EAI	-0.0046***	-0.0047***	VULN	0.022**		—
MPCE	—	-0.0001	1/sqEXP	—		-2.014***
FUEL	Failed to be added with $p < 0.10$	0.0257**	sqrtMPCE	Failed to be added with $p < 0.10$		Failed to be added with $p < 0.10$
Constant	-1.9775***	-2.1952***	Constant	-0.9864***		-0.7571***
LR $\chi^2(k)$	2030.46***	2107.32***	LR χ^2	51.17***		44.38***
Observations	131369	131369	Observations	9556		9556

*** Significant at less than 1% level.

** Significant at less than 5% level.

* Significant at less than 10% level.

In the following section, an account would be made about the factors in which the states differ and accordingly an attempt will be made to take out the homogenous group of states.

5.4. Disparities among Indian States and the Location of Disparity. The variations in the sociodemographic factors have substantial impact on the social and cultural practices that prevail in different parts of India. As a result, uneven development across the regions has been an integral feature of the country even after over sixty years of independence. Gini coefficient (The *Gini Coefficient* (also known as the *Gini index* or *Gini ratio*) is a measure of statistical dispersion intended to represent the income distribution of a nation's residents (Wikipedia).) of per capita Gross State Domestic Product (GSDP) has gone up consistently in the postreform period from 0.1917 in 1993-94 to 0.2409 in 2004-05 [5]. Percentage share of per capita GSDP, enjoyed by the top five states, has gone up from 28.3% in 1982-83 to 38.3% in 2000-01 and that of the bottom six states has been reduced from 35.3% to 26.9% [16]. Kundu and Varghese [5] have calculated the composite indices for economic conditions as well as the basic amenities and social dimensions which show the existence of the high inequality among the states of India. States in the higher position from the economic aspect cannot always provide better basic amenities and social development. Lack of basic amenities, along with economic backwardness, has a significant impact on both physical and mental health. Regional differences are, therefore, found in the morbidity and mortality patterns of the households of different states. Southern states are relatively healthier than the rest. A greater prevalence of illness and disability is observed in states like Uttar Pradesh, Bihar, and Madhya Pradesh. States of the southern region report their health to be good, whereas self-reported good health is very low in the case of Jammu and Kashmir, Jharkhand, and Assam. There are certain cultural and linguistic variations that can be observed in the propensity to respond spontaneously to the questions associated with status of health. Disparities are there among the states in terms of infant mortality rate as well as children and women's health. States differ with respect to the people's inclination to avail public medical care, hospitalization, travelling for treatment, and so forth [17]. Nair [18] presented a detailed analysis of the changing profile of the major Indian states with respect to both the economic development and social development in the postreform period. Definite tendencies towards regional divergence are noted in per capita NSDP in the Index of Infrastructural Development and in Human Development Index during the postreform period.

In the face of this diverse scenario, it would not be apt to analyse the impact of indoor pollution on the health-status of the population of India as a whole. For an effective intervention strategy, homogeneous states in terms of development indicators should be treated similarly. Attempts have been made in this section to group similar states to study the pattern of morbidity impact of air pollution more accurately. To identify the nature of interstate variation in this pattern, a two-step procedure has been applied here. In the first stage, *Multidimensional Scaling (MDS) Technique*

(Multidimensional scaling (MDS) refers to a broad class of procedures that scale objects based on a reduced set of new variables, derived from the original variables. It is specifically designed to provide a graphical representation of the objects in a multidimensional space (usually two or three-dimensional) such that the distances between the points in the space match the given dissimilarities as closely as possible. MDS is typically used as an exploratory method) is used on MPCE, EAI, and relative risk (RR) of acquiring morbidity through air-related diseases (Relative Risk (RR) is calculated as $RR = (\text{Probability of ABD in exposed group}) / (\text{Probability of ABD in unexposed group})$). Three basic pillars of development, namely, income, education, and health profile, are represented by these three variables. So, among others, they are considered to be the most appropriate basis for grouping the States/UTs in terms of their homogeneity. States with similar measures of MPCE, EAI, and RR are called homogeneous states and therefore clubbed together. They are homogeneous in the sense that they have similar income (MPCE), education (EAI), and health profile (RR). So, the difference in the impact of HAP in these states would not be influenced by these factors. In terms of policy, similar suggestions would be effective for homogeneous states. In the second stage, a cluster analysis is carried out at the state level on these variables in order to form different homogeneous groups.

5.5. MDS and States/UTs with Disparity. A classical MDS would provide a pictorial representation of the states closer to each other in terms of the Euclidean distances of the scores of the selected variables (i.e., MPCE, EAI, and RR). The objective is to reproduce a visual configuration of the distance matrix (matrix of pairwise distance of the scores between the states/UTs) in a small number of dimensions, so that the interpoint distances (fitted distances) are as close to the original distances as possible. (The technique has an interesting link with the Principal Component Analysis.) The closeness (goodness of fit) is judged by Kruskal's Stress Formula (Type I). (The formula aims to obtain the closest fit in the least square sense (classical scaling). The optimum configuration is determined by minimizing this measure of stress.)

Ratio MDS (In ratio scaling, the regression fit between interpoint distances in the configuration and the original distances goes through the origin) has been used in the present dataset. As mentioned, MDS would be applied on the average values of MPCE, EAI, and RR for 34 states/UTs except Nagaland, since the state did not report any victim of air-related disease and, therefore, the RR could not be defined. We had to take the mean as the average instead of the median in order to have a sufficient variation in the data set. Here, mean has appeared a more appropriate measure of the central tendency. We have data for the variables for 27 states and 7 Union Territories. So, a matrix of size 34×3 is generated. In the first stage of MDS, this 34×3 order data matrix is converted into a distance matrix, showing pairwise distances between states/UTs. Since the variables differ greatly in terms of their variances, they are first standardized to have a mean of 0 and a variance of 1 (standard normal variate).

Euclidean distances are computed. Since we apply ratio MDS fitted distances would be proportional to the original distances.

An MDS program looks for a spatial configuration of the objects, so that the distance between these objects match their proximities as closely as possible. The purpose of MDS of this dataset is to determine whether the states/UTs can be placed on a scale of development on the basis of these three indicators. One-dimensional solution, in that case, is of particular interest. This is presented in Figure 1 (SPSS version 16.0 has been used for running MDS (Alscal)). The problem of MDS is in understanding how to find a configuration of points that minimizes the squared differences between the optimally scaled proximities and the distances between the points. Thus, the goodness of fit of any MDS solution is indicated by stress measurement. The STRESS value (SPSS uses Kruskal type I for assessing the fit) for this single dimensional solution is 0.33, suggesting a poor fit (Kruskal and Wish [19, 20] have proposed assessment of fit using the following levels: $\text{STRESS} > .20$: Poor; $.10 \leq \text{STRESS} \leq .20$: Fair; $.05 \leq \text{STRESS} \leq .10$: Good; $.025 \leq \text{STRESS} \leq .05$: Excellent; $.00$: Perfect. Of course such straightforward interpretations should be handled flexibly, since STRESS is known to vary according to many other factors), because the reduction in a single dimension does not work well here. The same is indicated by RSQ (RSQ is the proportion of variance of the scaled data, which is accounted for by their corresponding distances) value ($\text{RSQ} = 0.79$) as well.

The iteration uses Young's S-STRESS formula (Young's S-STRESS formula is a measure of statistical fit. It ranges from 1 (indicating the worst possible fit) to 0 (indicating a perfect fit). It can be seen that there is an improvement (decrease) in Young's S-STRESS as the iterations proceed), and the solution needed 4 iterations, where improvement in Young's S-STRESS is less than 0.001. The figure shows that there are three clear outliers, Chandigarh and Dadra Nagar Haveli (DNH), lying at the upper extreme of the scale, and Lakshadweep (LD) at the lower end. However, the rest of the states, more or less, are concentrated around the centre. Before going into any further analysis, the scatterplot of this solution needs to be checked as we have already obtained a poor fit from the STRESS measurement. Figure 2 of the relevant scatter suggests a probe for a better fit with more dimensions.

Four iterations were required for the solution with two dimensions. The improvement in Young's S-STRESS is less than 0.001 for the final iteration. Kruskal's STRESS value for this solution is 0.12, which suggests a fair fit, an improvement from one-dimensional solution. The scatter (Figure 3) also reveals the same insight. In line with the one-dimensional solution, this configuration also separates Chandigarh, LD, and DNH, making them distinct outliers. In this two-dimensional representation, Bihar lies little far from the main clubbing of the states. (These states/UTs also stand out in a dendrogram when a hierarchical clustering method was applied to the dataset.) This is shown in Figure 4. The empirical findings on these three outliers explain their separate stand from the main concentration of states. First of all, it is worth noting that all the three outliers are Union Territories

FIGURE 1: Ratio MDS solution for one dimension.

FIGURE 2: Scatterplot of fitted and original distances for one-dimensional solution.

of India. The percentage of literates to the total population aged 7 years and above is 82 and 87 for Chandigarh and LD, respectively, while for DNH the figure is 57.6 only where the all India average is 64.8%. Moreover, in terms of certain infrastructure, like percentage of urban households with electricity in 2001 census, Chandigarh, DNH, and LD had 96.68, 95.84, and 99.67, respectively, where the all India average was 87.58. For the percentage of urban households with access to safe drinking water in the 2001 census, the figures for Chandigarh, DNH, and LD were 99.80, 96.10, and 4.60, respectively, whereas the all India average was as high as 90 [21]. So, these three UTs behaved differently in terms of literacy and infrastructure from the rest of the country, justifying their position in the MDS analysis as described here.

To get a precise idea about how the states/UTs are grouped on the basis of their development indicators, a cluster analysis has been taken up as well, where the grouping of the states/UTs is distinctly identified. Clustering of 34 states/UTs would provide another check to see whether the outliers obtained from the MDS solution are behaving separately or not.

5.6. Clustering and Grouping of States/UTs. This initial analysis of MDS suggests the presence of certain wide disparities

FIGURE 3: Scatterplot of fitted and original distances for two-dimensional solution.

FIGURE 4: Ratio MDS solution for two dimensions.

between Indian states and therefore, to study the latent characteristics of these different groups of states in greater details, a *hierarchical* cluster analysis has been carried out next, where the clustering is done by following the *between group* linkage method. This is carried out on the basis of the same three variables, MPCE, RR, and EAI. Variables are taken in their standard normal form to bring them on the same scale of variability.

The visual approximation from two-dimensional MDS plot does not provide any clear picture regarding the number of clusters, though the observations are spread over four quadrants of the two-dimensional plane. The dendrogram (Figure 5) obtained from the hierarchical cluster method provides 7 groups with six singletons, namely, Uttarakhand, Jharkhand, Bihar, Chandigarh, DNH, and LD. The last three confirm their outlying behavior as obtained in the MDS analysis. The first seven clusters are clubbed and reduced to three clusters on the basis of the distances between the groups. Thus, the first group includes Sikkim, Mizoram, Himachal Pradesh, Gujarat, Punjab, Andhra Pradesh, Daman and Diu, Andaman and Nicobar, Orissa, Madhya Pradesh, Uttar Pradesh, Tripura, Tamil Nadu, Jammu and Kashmir, Rajasthan, Karnataka, Pondicherry, and Goa (a total of 5607 observations); the second group is comprised of Arunachal Pradesh, Manipur, Assam, Chhattisgarh, West

Bengal, Meghalaya, Kerala, Uttarakhand, and Jharkhand (2143 observations). In spite of their separate standing, Uttarakhand and Jharkhand are included into this cluster as they are positioned nearby. In the third cluster, Haryana, Maharashtra, and Delhi are clubbed together (with a total of 1438 observations). The states in a particular cluster have a homogeneous infrastructure in terms of health, education, and income. In fact, it is explained by the technique through which they have been clubbed. The three outliers, along with Bihar, are dropped from the subsequent analysis as they stand considerably further from the other three clusters.

In Table 6, the averages of Net State Domestic Product (NSDP) for the states/UTs in three clusters are reported. The percentage of population that uses dirty fuel in the urban areas, incidence of air-related diseases, and the relative risk of being affected by air-induced diseases among the dirty fuel users are reported for the three clusters separately. Considerable differences are found in all the variables, implying the disparate nature of the states in separate clusters in terms of fuel type, health status, and income level.

5.7. Exploratory Analysis for Clusters. Data exploration in each of the clusters was carried out, following the same method described earlier for the all India analysis. MPCE is dropped in this analysis, as discussed before, without considering the ambient effect it has and is unable to explain the air-related morbidity as a function of HAP in the all India case. The descriptive statistics for each cluster are reported separately in Table 7.

Barring some exceptions, the behavior of variables across the clusters is mostly uniform. In all the clusters, transformation is needed for EXP only. EXP, with a mean value higher than median and with higher SD than pseudo-SD, needed an inverse transformation for the second cluster ($1/EXP$) and an inverse square transformation ($1/\sqrt{EXP}$) for the other two. The ladder analysis failed to suggest any significant improvement of goodness of fit with any other functional form for the EAI. Therefore, it has been left in its original (raw) form. To illustrate the improvements attained through variable transformations, we have also checked the box plots of this variable for a few relevant clusters. This is shown in (Figure 8). The transformation of the variables is carried out to obtain a symmetric (normal or close to normal) distributions required for running a standard regression analysis.

In the next section, an attempt would be made to analyse the relationship between morbidity and HAP, giving control to other socioeconomic influences for each cluster separately and the results would be then compared.

5.8. Morbidity Relationships in Clusters. The variables mentioned in the third section would be used for the analysis. ABD is the dependent variable, and the influence of other variables on ABD will be examined after carrying out the required transformation, as mentioned in the previous section. All the regressions would be at the individual level. Any presence of multicollinearity has been checked through

FIGURE 5: Dendrogram using average linkage (between groups).

FIGURE 6: Health-damaging pollutants per unit energy delivered: ratio of emissions to LPG. Log scale is used in the figure. The values are shown as grams per megajoule of energy delivered to the cooking pot (g/MJ-d).

Variance Inflationary Factor (VIF), and no compelling threat has been detected for any of the clusters.

Like the all India analysis, stepwise regression has been conducted in three specific clusters. The results are shown for each cluster separately. The addition of the variables into the model has been restricted to 10% significance level only.

For each cluster, two sets of regressions were run. The first set takes EMSN and VULN separately, and the second set includes EXP (in its appropriate form) instead of the former two. The results are reported in Table 8.

From this table, a few observations can be made:
(i) LR Chi square is significant in all regressions, confirming their statistical acceptability.

(ii) For cluster 1, EAI is the most important factor, which explains the incidence of air-related morbidity. Emission of particulate matters from different fuels has a significant contribution in the incidence of ABD, confirming the hypothesis that HAP generated from fuel emission has a hazardous effect on human health. When exposure is used instead of emission and vulnerability (R₆), it has a significant impact on the air-induced morbidity at the individual level. In R₆, education has a significant impact on ABD. It extends support to the fact that air-related morbidity can be addressed by improving the education attainment, which would indirectly help in altering the behavioral pattern of the individual by improving awareness to take up measures for combating the HAP problem. In fact, an increased awareness can improve the living condition of the household, which can be most helpful in controlling the problem in a magnified manner.

(iii) 41% of the urban residents of the states in cluster 2 use dirty fuels and the relative risk of getting affected by air-related morbidity is as high as 89% (Table 6). It leads to an anticipation, where exposure should have a direct impact on the incidence of ABD for this cluster. This is reflected in

TABLE 6: Clusterwise variation for selected variables[#].

Clusters	Per capita NSDP ^{##} (in Rupees at 2004-05 prices)		Dirty fuel user	Incidence of air-related disease	Relative risk for ABD
1	Mean	1742.529	30%	18%	53%
	CV*	3.39			
2	Mean	9670.556	41%	17%	89%
	CV	1.20			
3	Mean	460.00	11%	20%	16%
	CV	0.34			

* Coefficient of variation.

Source: [22].

[#]NSDP is recorded from a different data source; FUEL and M were not chosen for the clustering done here earlier.

^{##}In Cluster 1, the average NSDP does not include the data for Daman and Diu as the information was not available in the data source.

the results from R_8, where exposure is included first prior to EAI and becomes significant with expected sign (negative of 1/EXP means a positive relation with EXP). Emission plays an important role here (R_7) in explaining the incidences of this illness. Educational attainment is important for this group of states, which show that a significant leap in the education level can handle the situation better.

(iv) For cluster 3, the exposure only becomes significant (R_10). This suggests that the reduction in the level of individual exposure can have a substantial influence on the incidence of air-related diseases in this cluster. Vulnerability, based on age and sex, also becomes significant when added in the model individually confirming the fact that same level of pollution affects different individuals in different ways mainly on the basis of their power of resistance to the diseases as reflected in the age-sex profile of the person concerned.

(v) Individual level incidence of morbidity in the states in cluster 1 and 2 is likely to respond to a policy initiative on enhancement of educational attainment. An inspection to the coefficients of the variable in cluster 1 and 2 suggests that the incidence would be marginally more responsive to the improvement in the education level for the group of states in cluster 2. But bringing any change in the education and awareness needs a long-term and thoughtful planning on the part of the state.

(vi) A shift in the fuel use would improve the exposure and hence the incidence of this particular morbidity in all three clusters may be seen. Benefits would be substantial for all the states.

Finally, on the basis of the results in Table 8, a policy suggestion can be framed for the different clusters. A brief account is made in Table 9 concerned with the effective policy instruments that are left separate for each cluster. This is a summarization of the results of the previous analysis. EAI, EMSN, and EXP are chosen as the appropriate variables for taking up any policy initiative. VULN has not been considered to be a policy variable as this depends on age and sex of the individual which cannot be changed. Effective policy instruments are discussed clusterwise.

(vii) For clusters 1 and 2, educational improvement and reduction in the emission would improve the incidences of morbidity. Emission inside the house can be improved by

carrying out a shift in the fuel used by the household. Better and safer fuel replacement is possible if proper subsidization scheme on the part of the government can be taken up, although there are several other factors that may hinder the success of this initiative as discussed previously. For these states, the exposure reduction could be an important policy option. Again, better fuel choice becomes important for any change in the exposure level.

It is clear from this study that the problem of HAP has a social dimension and should be handled from a socio-economic perspective. Barnes [23] showed that a comprehensive Chinese behavioral trial tested the effectiveness of "health education and behavioral activities" (HEBA) together with improved cook stoves in four rural provinces of China (Gansu, Guizhou, Shaanxi, and Inner Mongolia). The combination of HEBA with improved stoves showed measurable improvements in the indoor air quality (by as much as 85%) and efficiency. A study by Johnson et al. [24], concerning simple low-cost interventions, directed to producing cleaner indoor air, coupled with healthy home education, improved the indoor air quality and the health of asthmatic children. Parikh et al. [25], Budds et al. [26], and Edelstein et al. [27] mentioned that education and cultural modifications can be effective in increasing the awareness of the health effects of HAP and hence reduce the exposure to the same [28]. So, educational improvement in the form of provision of primary level education at the initial stage, HAP awareness program, and so forth can help in improving the situation and reducing morbidity.

The most popular policy target could be an adoption of slum development program, which is likely to have a positive effect on indoor pollution. Some such programs were mentioned by Dutta [29]. The study referred to some of the programmes which were taken mainly by the Kolkata Metropolitan Corporation (KMC) for slum improvement. The needs and problems of the slum dwellers in Kolkata Metropolitan Corporation (KMC) have been seriously considered by both KMC and Kolkata Metropolitan Development Authority (KMDA) for decades. Since late '80s and early '90s, KMC has started to take initiatives for the improvement of the condition of the slum dwellers. The *Calcutta Slum Improvement Project* (CSIP) is an initiative in consultation with the Urban

TABLE 7: Descriptive statistics for variables in different clusters.

Variables	Cluster 1	Cluster 2	Cluster 3
EXP			
Mean	60.60	62.22	27.60
Median	8.00	9.00	6.00
SD	97.17	108.28	73.87
Pseudo-SD	92.66	91.92	2.96
Skewness	2.40	3.40	5.72
Kurtosis (normalized)	8.56	16.76	45.52
Required transformation	1/sq	Inverse	1/sq
EAI			
Mean	46.10	55.67	55.15
Median	50.00	66.67	66.67
SD	37.40	34.09	35.84
Pseudo-SD	61.77	37.06	37.06
Skewness	0.0075	-0.34	-0.36
Kurtosis (normalized)	-1.44	-1.05	-1.20
Required transformation	Raw	Raw	Raw
VULN			
Mean	6.36	6.47	6.37
Median	5.00	6.00	5.00
SD*	1.64	1.67	1.65
Pseudo-SD	2.22	2.22	2.22
Skewness	0.63	0.51	0.63
Kurtosis (normalized)	-1.37	-1.51	-1.37
Required transformation	Takes certain values, no transformation	Takes certain values, no transformation	Takes certain values, no transformation
EMSN			
Mean	9.54	9.81	4.39
Median	1.00	1.00	1.00
SD*	14.71	16.22	11.19
Pseudo-SD	18.53	18.53	0
Skewness	2.09	2.87	4.87
Kurtosis (normalized)	5.92	11.16	30.84
Required transformation	Takes certain values, no transformation	Takes certain values, no transformation	Takes certain values, no transformation

*SD: standard deviation. Source: author's calculation.

Poverty Office of the DFID of the Government of UK. The main objectives of the programme included the improvement of unhygienic environment of slums through improvement of sanitization model of the infrastructure, provision of primary health care services at the doorstep of the slum dwellers, and additionally providing family welfare, nutrition program, educational awareness, and improved economic condition through income generation programmes, and so forth. Interventions were in the form of portable water supply, sanitary latrines, effective drainage system, and establishment of health infrastructure as well as economic and educational support program. Though these policies did not target the reduction of HAP directly, when income and other conditions improved, the representative household moved up in

the energy ladder and went for a cleaner fuel choice. Awareness of the adverse health outcomes of pollution exposure went up, and the affordability of safer sources of fuel increased. The combined effect led to a reduction in the exposure burden of the most vulnerable group, slum women, children, the elderly, and the unemployed. But all these programs needed a long period of time in order to be implemented. This is an eventual policy instrument to be effective in long run.

(i) Exposure is the only area where policy intervention is possible for cluster 3.

There is a gender perspective on this issue of exposure to HAP. The burden mainly falls on women and children. The study by Dutta and Banerjee [30] showed that women

TABLE 8: Stepwise probit regression results for HAP and morbidity: three clusters.

	Cluster 1			Cluster 2			Cluster 3		
	(Dependent variable ABD)			(Dependent variable ABD)			(Dependent variable ABD)		
R.5	$p = 0.0000 < 0.1000$ adding EAI			$p = 0.0015 < 0.1000$ adding EAI			$p = 0.0735 < 0.1000$ adding VULN		
	$p = 0.0159 < 0.1000$ adding EMSN			$p = 0.0191 < 0.1000$ adding EMSN					
R.6	$p = 0.0000 < 0.1000$ adding EAI			$p = 0.0004 < 0.1000$ adding I/EXP			$p = 0.0378 < 0.1000$ adding I/sqEXP		
	$p = 0.0603 < 0.1000$ adding I/sqEXP			$p = 0.0480 < 0.1000$ adding EAI					
R.5	R.6			R.7	R.8		R.9	R.10	
EAI	-0.0020***	EAI	-0.0021***	EAI	-0.0023***	I/EXP	-1.1566**	VULN	0.0404*
EMSN	0.0032**	I/sqEXP	-2.3955*	EMSN	0.0045**	EAI	-0.0020**	EAI	Failed to be added with $p < 0.10$
VULN	Failed to be added with $p < 0.10$	—	—	VULN	Failed to be added with $p < 0.10$	—	—	EMSN	Failed to be added with $p < 0.10$
Constant	-0.8601***			Constant	-0.8833***	Constant	-0.7385***	Constant	-1.095***
LR χ^2	27.49***			LR χ^2	15.45***	LR χ^2	16.34***	LR χ^2	3.19*
Observation	5607	5607	5607	Observation	2143	Observation	2143	Observation	1438

*** Significant at less than 1% level; ** significant at less than 5% level; * significant at less than 10% level.

TABLE 9: Policy suggestion for each cluster.

Clusters	Policy instrument Incidence of morbidity		
	EAI	EMSN	EXP
1	Effective	Effective	Effective
2	Effective	Effective	Effective
3	Not effective	Not effective	Effective

Source: prepared on the basis of the previous analysis.

in urban India are more affected by air-related diseases, particularly in the age group 14–40, when they participate in cooking activities. Within each household, the risk of the females is much higher than the males. The study suggested that the freedom of women to choose the fuel type, with proper awareness, could provide a solution to the problem.

6. Conclusion

The analysis establishes that HAP, in terms of fuel use, explains the general morbidity for the urban individuals of India. Additionally, fuel emission and exposure at the individual level have an impact on air-related diseases. But the income level of the individual fails to explain the incidence of air-related morbidity, implying a presence of residual effects. Indoor and outdoor pollutions both cause air-related diseases, and it is not a sole outcome of HAP. This explains the absence of any relation between ABD and income in the analysis. The income has been dropped as a policy option in the rest of the analysis. This could be an area for future research, taking ambient pollution into consideration while developing such an analysis. Additionally, an explanation for income improvement and the increase in vulnerability associated with exposure can be provided if qualitative data at the primary level are made available.

States/UTs have been clubbed with respect to their developmental indicators so that similar policies for combating HAP can be taken for those who stand in the same line of development. Factors like awareness and exposure have been identified as the most appropriate policy areas. In a diverse country like India, the states differ in opinion with respect to each of these factors. So, designing any single policy for the whole country, based on these factors, is both complex and irrational. An attempt has been made in this analysis to ease out the policy strategies by locating the states with similar infrastructural background.

States with homogeneous structure have been grouped together and policies are suggested separately for each group. The suggestions include the change in attainment of education and switching the fuel type, though the nature of the policy differs across groups. With regard to the fuel choice, complete fuel switching is not possible as most of the households maintain a portfolio of energy sources. A detailed study on the portfolio pattern of fuel used by the households and the factors on which this pattern depends upon could be an agenda for future research. Moreover, the research can be carried out to look into the impact of different portfolios

about the morbidity pattern. A primary level data, collected at the source of the pollution, can be effective in capturing the impact of HAP on air-related morbidity in a more accurate manner. Data should be provided about the actual concentration of pollutants coming out from indoor cooking mediums and other sources, ventilation structure, location of the kitchen, open space in the house, congestion, and so forth for a more comprehensive analysis of the matter. Any subsidization scheme provided till date has not been marked as an effective policy strategy. At another point, the problem of HAP can be handled from the point of bringing effective changes in the living condition such as by making changes in the in-house ventilation system, installation of exhaust fans, creating an arrangement for building partitions in the homes to separate cooking and sleeping areas, and improvement in the subsidiary living condition parameters like drainage, latrine, and house structure (which might have a secondary effect on HAP-related morbidity by altering the congested structure of the urban houses especially slums). Any change in the living condition is possible only in the presence of sufficient affordability on the part of the individual. For that, analysis of the effects on HAP including income in the study has to be made. This is an important part of the research and without the information on ambient pollution the impact on air-related disease is both incomplete and inadequate. A considerable improvement in the income level, which would lead to the person being able to make changes in the living condition, can be a commendable solution. But this, to a large extent, comes under the public domain. In fact, changing the income level of the individual needs an improved and broad long-run planning from the grass-root level and also relates to the sociopolitical environment of the nation. Behavioral pattern of the household members is also important in the context of effective policy design. So, alterations in the behavioral pattern of the household should be the first priority of any strategic intervention program. Educational training becomes important in this context. As we have obtained from our analysis, educational attainment has been important for the first two groups of states, where mere literacy might not be sufficient for the realization of the required improvement in health. Education up to primary level, or even more, is necessary to conceptualize the hazardous effects of HAP on the morbidity level of the household. A targeted training program on health and hygiene, at the school level, may work well in this context. Since educational policy is a state subject and does not depend much on individual decision, it should be the first priority to be pinned upon. The word of hope is that, in the last few years, the spread of elementary education has improved significantly. Under Sarva Shiksha Abhiyan, the Gross Enrolment Ratio in both primary and upper primary levels has increased and the dropout rates from both levels have reduced [31], although the rates differ across the states. Changing one's perception is challenging and needs sufficient time to become reality. Once people are aware and convinced of the problems, then the policy benefits look more obvious. India, with its huge disparity in structural composition, would be able to enjoy better health if the policies are taken on the basis of the specified features of the individual state. A single policy for the entire country

FIGURE 7: GLADDER graph for MPCE. Histograms by transformation.

FIGURE 8: Box plot and transformation of selected variables.

is unlikely to work in reducing the indoor pollution-induced morbidity and improving health in general. An integrated and well-designed educational policy, with a special attention to health and hygiene, may hold the problem at its root, and an overall change in the fuel-use pattern, living conditions, and the behavior of the victims could be brought forward as the immediate outcomes.

Appendix

Data Exploration and Variable Transformation. If the mean is larger than the median, then the distributions are positively skewed. On the basis of the Inter Quartile Range (IQR), a pseudo-SD is estimated. If this is smaller than the standard deviation, then the tails are heavier than normal. Thus, an idea about the required transformation of data can be formed by comparing mean-based vis-à-vis order-based statistics of the data when one is trying to create a model of the average behavior [32]. Generally, variables with fat tail and positive skewness require a square root or logarithmic transformation to pull in outliers from the higher end. Sometimes, an inverse of the variable or of the squared or square root values seems to be the fittest transformation. For negatively skewed distribution with the fat left tail (i.e., outliers at the lower end), power transformation X^p where p equals to 2 works better. The required transformations are reported in Table 2. Here,

we have used the GLADDER technique in STATA to derive the appropriate transformation. It uses a graphical approach which is supported by χ^2 goodness of fit (GLADDER graph for MPCE only is given in Figure 7).

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The present work immensely benefitted from Professor Sarmila Banerjee and Rajiv Gandhi, Chair Professor, Department of Economics, Calcutta University, who provided her valuable comments, ideas, and assistance to the writing and undertaking of the research summarized here. Discussions with Dr. Mousumi Dutta, Associate Professor, Presidency University, Kolkata, have further helped in shaping the ideas in a better manner.

References

- [1] World Health Organization, *Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks*, World Health Organization, 2009, <http://www.who.int/>.

- [2] N. El-Hougeiri and M. El Fadel, "Correlation of indoor-outdoor air quality in urban areas," *Indoor and Built Environment*, vol. 13, no. 6, pp. 421–431, 2004.
- [3] S. Ishido, *Variations in Indoor and Outdoor Dust Densities*, Osaka City University, Osaka, Japan, 1955.
- [4] D. G. Fullerton, N. Bruce, and S. B. Gordona, "Household air pollution from biomass fuel smoke is a major health concern in the developing world," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 102, pp. 843–851, 2008.
- [5] A. Kundu and K. Varghese, *Regional Inequality and Inclusive Growth in India under Globalisation: Identification for Lagging States for Strategic Intervention*, vol. 6 of *Oxfam India Working Papers*, Oxfam India, 2010.
- [6] S. Sankhe, I. Vittal, R. Dobbs et al., "India's urban awakening: building inclusive cities, sustaining economic growth," McKinsey Global Institute Report, 2010.
- [7] IIPS, "National family health survey," Report 3, 2005–2006.
- [8] B. S. Reddy, "Economic and social dimensions of household energy use: a case study of India," in *Proceedings of the 4th Biennial International Workshop on 'Advances in Energy Studies'*, E. Ortega and S. Ulgiati, Eds., pp. 469–477, UNICAMP, Campinas, Brazil, 2004.
- [9] K. Smith, J. Rogers, and S. C. Cowlin, *Household Fuels and Ill-Health in Developing Countries—What Improvements Can be Brought by LP Gas?* World LP Gas Association, Paris, France, 2005.
- [10] R. Smith Kirk, "Household air pollution and acute respiratory infections in editorial," *Indian Paediatrics*, vol. 40, pp. 815–819, 2003.
- [11] K. R. Smith, J. M. Samet, I. Romieu, and N. Bruce, "Indoor air pollution in developing countries and acute lower respiratory infections in children," *Thorax*, vol. 55, no. 6, pp. 518–532, 2000.
- [12] K. R. Smith, "National burden of disease in India from indoor air pollution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 24, pp. 13286–13293, 2000.
- [13] V. Kathuria and N. A. Khan, "Vulnerability to air pollution: is there any inequality in exposure?" *Economic and Political Weekly*, vol. 42, pp. 3158–3165, 2007.
- [14] D. Y. C. Leung, "Outdoor-indoor air pollution in urban environment: challenges and opportunity," *Frontiers in Environmental Science*, vol. 2, article 69, 7 pages, 2015.
- [15] M. Ezzati and D. M. Kammen, "The health impacts of exposure to indoor air pollution from solid fuels in developing countries: Knowledge, gaps, and data needs," *Environmental Health Perspectives*, vol. 110, no. 11, pp. 1057–1068, 2002.
- [16] S. L. Shetty, "Growth of SDP and structural changes in state economies: interstate comparisons," *Economic and Political Weekly*, vol. 38, no. 49, pp. 5189–5200, 2003.
- [17] S. B. Desai, A. Dubey, B. Joshi, M. Sen, A. Shariff, and R. D. Vanneman, *Human Development in India: Challenges for A Society in Transition*, Oxford University Press, Oxford, UK, 2010.
- [18] K. R. G. Nair, "Economic Reforms and Regional Disparities in Economic and Social Development In India (Report of a Research Project funded by the SER Division of the Planning Commission of the Government of India)," 2004, <http://www.planningcommission.nic.in/>.
- [19] D. J. Bartholomew, F. Steele, I. Moustaki, and J. I. Galbraith, "Multidimensional scaling," in *Analysis of Multivariate Social Science Data*, chapter 3, Chapman & Hall, 2nd edition, 2008, <http://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/chapter3.pdf>.
- [20] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.
- [21] <http://www.cbhi.nic.in/>.
- [22] Reserve Bank of India, *Handbook of Statistics on Indian Economy*, Reserve Bank of India, 2011–2012.
- [23] B. R. Barnes, "Behavioural change, indoor air pollution and child respiratory health in developing countries: a review," *International Journal of Environmental Research and Public Health*, vol. 11, no. 5, pp. 4607–4618, 2014.
- [24] L. Johnson, C. Ciaccio, C. S. Barnes et al., "Low-cost interventions improve indoor air quality and children's health," *Allergy and Asthma Proceedings*, vol. 30, no. 4, pp. 377–385, 2009.
- [25] J. Parikh, K. Parikh, and V. L. Pandey, "Lack of energy, water and sanitation and its impact on rural India," in *India Development Report, 2004–2005*, K. Parikh, Ed., pp. 84–95, Oxford University Press, Oxford, UK, 2005.
- [26] J. Budds, A. Biran, and J. Rouse, "What's cooking? A review of the healthy impacts of Household air pollution and technical interventions for its reduction," WELL Task 512, Loughborough University, Water, Engineering and Development Centre, Leicestershire, UK, 2001.
- [27] M. Edelstein, E. Pitchforth, G. Asres, M. Silverman, and N. Kulkarni, "Awareness of health effects of cooking smoke among women in the Gondar Region of Ethiopia: a pilot survey," *BMC International Health and Human Rights*, vol. 8, article 10, 2008.
- [28] G. Hu and P. Ran, "Indoor air pollution as a lung health hazard: focus on populous countries," *Current Opinion in Pulmonary Medicine*, vol. 15, no. 2, pp. 158–164, 2009.
- [29] S. Dutta, "Air pollution exposure in Kolkata: an assessment of decadal change," *Artha Beekshan*, vol. 17, no. 3, 2008.
- [30] S. Dutta and S. Banerjee, "Exposure to household air pollution and women health: the situation in urban India," *Environment and Urbanisation ASIA*, vol. 5, no. 1, 2014.
- [31] Ministry of Human Resource and Development, *Working Group Report on Elementary Education and Literacy*, 2011, <http://planningcommission.nic.in>.
- [32] C. Mukherjee, H. White, and M. Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, London, UK, 1998.