Urban-rural differences in infant mortality in India

Aashish Gupta*

September 19, 2018

Abstract

Rural areas in contemporary developing countries have higher infant mortality. This is in contrast with historical patterns in developed countries, where urban areas faced a mortality penalty. This paper investigates rural-urban differences in infant mortality in India, home to one-fifth of all infant deaths, using newly released Demographic and Health Survey data. It examines the extent to which environmental exposures, socioeconomic status, and access to health can explain higher child survival in urban areas. Using two complementary strategies, regression adjustment and Oaxaca-Blinder decomposition, I find that environmental exposures, and in particular, solid fuel use, can explain a large portion of the difference in infant mortality between rural and urban areas. In contrast, rural disadvantage in terms of individual, household, or community-level socioeconomic status or access to health explains a much smaller part of the urban-rural difference. I discuss the theoretical and policy implications of this surprising finding. From a policy perspective, it is important not to think of pollution as only an urban problem. Similarly, efforts to improve health in rural areas should not be limited to improving health access. Theoretically, these findings have important implications for understanding the ongoing epidemiological transition in developing countries. The findings indicate that that although patterns of rural-urban mortality disadvantage are reversed in contemporary developing countries, much like historical developed countries, environmental exposures play an important role in generating these differences. Finally, I outline areas of further work which will strengthen the analysis in this paper. (243 words)

`Keywords: infant mortality, development, urbanization, epidemiological transition, environment, solid fuels, India`

---

*Email: aashishg@sas.upenn.edu. Ph.D. Student, Demography & Sociology, University of Pennsylvania. I am grateful to Jere Berhman, Diane Coffey, Jean Dreze, Irma Elo, Michel Guillot, Devesh Kapur, Kanika Sharma, Dean Spears, Atheendar Venkataramani, and Anumeha Yadav for helpful comments and suggestions. All errors are my own.

1 Introduction

Understanding the relationship between place and health has been a central concern within population, health, and environmental sciences (Hippocrates, 2004; Snow, 1856; Macintyre et al., 2002). Importantly, a long tradition within this literature has inquired into rural-urban differences in health, investigating the causes of urban mortality penalty (Condran and Crimmins, 1980; Haines, 2001) or rural disadvantage (Poel et al., 2009; Pörtner and Su, 2018; Sastry, 1997), while highlighting policy approaches that addressed or would address disadvantages (Preston and Haines, 1991; Cutler and Miller, 2005).

Investigations of differences where historical data is available, such as in contemporary developed countries, have found a historical urban mortality penalty (Hubbard, 2000; Reher, 2001), uncovering the health costs of the industrial revolution (Beach and Hanlon, 2016; Kesztenbaum and Rosenthal, 2011). On the other hand, studies in contemporary developing countries, based primarily on Demographic and Health Surveys (DHS) data, have found persistently worse child health in rural areas (Poel et al., 2007; Bocquier et al., 2011).

Because rural areas in developing countries are poorer and socioeconomically disadvantaged (Chen and Ravallion, 2007), studies have examined the extent to which socioeconomic disadvantages can explain health disparities between rural and urban areas (Sastry, 1997; Saikia et al., 2013). However, the role of environmental exposures, such as sanitation and pollution, which have been found to explain historical urban penalties in urban areas of developed countries (Condran and Crimmins, 1980; Beach and Hanlon, 2016), and exposure to which is higher in rural areas of current developing country contexts (World Health Organization, 2015; Rehfuess et al., 2006), has been less studied in the context of understanding rural disadvantages. The paper aims to fill this research gap for the important case of India.

A substantial portion of the world's infant deaths occur in India, urbanization rates are low, and rural-urban disparities are large. Out of 4,242,000 total infant deaths in 2016, India was home to 867,000 infant deaths, or more than one in five infant deaths (You et al., 2017).

It also had the highest number of infant deaths in any country. 68.9% of the population lived in rural areas (Government of India, 2011). Rural infant mortality rates in the five years preceding the latest DHS in India (National Family Health Survey-4 2015-16) were estimated to be 45.5 per 1,000 live births, 1.6 times that of urban areas.

Prior research on rural-urban disparities in child health in India has found that in the earlier rounds of the DHS (1992-93, 1998-99, 2005-06), accounting for the DHS wealth-index, maternal characteristics (age and education), and other health-care related measures (vaccination and place of delivery) could explain a substantial portion of the difference (Saikia et al., 2013). Poel et al. (2007) finds similar results for rural disadvantage in childhood stunting. As with other research on health gradients (Adler et al., 1993; Elo, 2009), the mechanisms through which differences in household wealth or maternal education explain rural disadvantages in infant mortality need to be considered, but have not yet been investigated. This is despite the clear role of uncovering mechanisms in classical frameworks of understanding social gradients in child mortality, such as those of Mosley and Chen (1984). Additionally, the standardized DHS wealth index includes household use of toilets or cleaner fuel for cooking, which contribute to improving the disease environment, while also being indicators of household wealth. Thus, controlling for DHS wealth index also controls for environmental exposures to some extent. Finally, improvements in health-services in rural areas and rural poverty declines warrant a re-examination of the factors contributing to continuing rural infant mortality disadvantage.

The paper uses several strategies to investigate rural-urban differences. First, we use two complementary methods to understand why rural areas have higher infant mortality - multivariate OLS regressions and Oaxaca-Blinder decomposition methods. Second, to address the role of environmental exposures and externalities, I estimate the proportion of other households living near a child that use solid fuels or which defecate in the open. This approach has been used in prior research using DHS surveys to estimate the impact of neighborhood sanitation (Geruso and Spears, 2018) and vaccination coverage (McGovern and Canning, 2015) on child mortality, but has not been used in the context of solid fuel use, which contributes to local air pollution (Chafe et al., 2014) and harms children's health (Smith et al., 2011). This approach can also help resolve why within the extant empirical literature, solid fuel use "accounts for fewer child deaths than suggested by earlier indirect estimates" (Bassani et al., 2010), highlighting that children in comparison households may also be exposed to smoke because of solid fuel externalities. Third, to avoid measurement errors, I limit the estimation sample to live births 24 months before the date of the survey. Within the literature, it is common to include births 5 years before the date of interview, however, it is likely that environmental exposures or household wealth were different 5 years before the date of interview. Finally, I construct a principal components analysis (PCA) based wealth index using the same methods as used by the DHS, but include only those assets that capture household wealth and are not directly linked to child health (an example of an asset directly linked to solid fuel use would be clean fuel use).

Further details on the data and methods used are provided in section 2. Section 3 presents results, while section 5 discusses the implications of the findings. I provide an outline for further work that would strengthen the paper in section 4.

2 Data and methods

2.1 Data

The data used in this study come from the birth history module of the latest DHS in India (NFHS-4 2015-16). NFHS-4 was representative at the national, state, and district levels, and like other standard DHS surveys, interviewed women aged 15-49 about children ever born and their mortality in 28,522 clusters; 22 households were selected from each cluster or primary sampling unit. Overall sample size was 601,509 households, with a 98% response rate (IIPS & ICF, 2017). From a total of 699,686 female respondents, information on 1,315,617 births was collected. For the analysis here, the sample has been restricted to births between

12 and 24 months before the survey. This restriction allows every child to be exposed to 12 months of risk of dying. It also avoids measurement error, because household characteristics would have likely been different for children born earlier than 24 months. This way, the final analytical sample is restricted to 54,722 live births, born between January 2013 and November 2015. Descriptive statistics for the sample are available in Table 1. Throughout the analysis, survey weights are used and standard errors are clustered at the level of the primary sampling unit.

Additionally, using the geographic coordinate locations of DHS clusters and birth year of the child, I matched satellite derived annual pollution estimates by Van Donkelaar et al. (2016) to the analytic sample. Appendix Figure A1 shows global annual estimates of particulate matter, calculated as particulate matter less than 2.5 microns per cubic meter ($\mu g/m^3$) for the year 2014 from these data. Appendix Figure A2 shows matched DHS clusters with pollution estimates.

2.1.1 Outcome

The main outcome used in the study is Infant mortality, and we are interested in urban-rural differences in IMR. Table 1 shows that for the analytical sample used here, overall Infant Mortality was 40.3 deaths per 1,000 live births. Infant mortality in urban areas was lower, as expected, with 25.2 deaths per 1,000 live births. IMR in rural areas was 46.1 per 1,000. These overall rates are similar to reported DHS/NFHS summary statistics (IIPS & ICF, 2017).

2.1.2 Exposures & covariates

We adjust for demographic covariates of the child, and are interested in the extent to which maternal characteristics, measures of socioeconomic status, and environmental exposures can explain the rural infant mortality disadvantage.

Demographic characteristics: Covariates adjusted for in the models are dummies for year of birth of the child, dummies for month of birth of the child, dummies for birth order, and sex of the child. Live births were equally likely to be male than female, but were not more male-biased in urban areas when compared to rural areas. Birth order was slightly higher in rural areas, reflecting higher fertility.

Maternal characteristics: Within the literature, there has been a long discussion of the role of maternal education in reducing child mortality (Caldwell and McDonald, 1982). We adjust for maternal characteristics, such as maternal age, maternal bmi, and mother's education. Mothers were slightly older in urban areas, and maternal bmi and education were also higher in urban areas.

Measures of socioeconomic status: Here we are concerned with those measures which are linked to child health less directly. First, I created a asset score based on principal components analysis. The assets included to create this score were household ownership of radio, television, refrigerator, bicycle, motor-cycle, car, land-line phone, mobile phone, watch, bank account, mattress, pressure cooker, bed, table, fan, sewing machine, internet connection, computer, air cooler, and washing maching. I took the first component of the analysis, and for all the assets, the Kaiser-Meyer-Olkin measure of sampling adequacy was above 0.5, justifying inclusion in the model. As expected, the asset score was higher in urban areas than in rural areas. I also calculated the average asset score of other households in a child's primary sampling unit, which was also higher in urban areas. Household and neighborhood electrification was high in both urban and rural areas, though higher in urban areas.

Access to health: In this version, we use the place of delivery of the child, and the person who conducted the delivery. Children born in rural areas were less likely to have been delivered in a health facility, or to have received the attention of a health worker. Only a very small proportion were at births without the assistance of a doctor or health worker.

Environmental exposures: In this analysis, we use two kinds of environmental exposures: exposure to fecal matter, as proxied by neighborhood open defecation, and household and ambient air pollution, as measured by household and neighborhood solid fuel use. Additionally, we use satellite derived PM 2.5 levels as another measure of ambient air pollution.

2.2 Empirical strategy

The empirical strategy consists of examining descriptive relationships between the exposures and rural-urban differences in infant mortality, a multivariate regression approach, and Oaxaca-Blinder decompositions. The estimated ordinary least squares regression equation is as follows:

$$\begin{aligned} infant\,mortality_{ij} =& \beta_1 rural_{ij} + \beta_2 explanatary\,factors_{ij} \\ & + \theta demographic\,controls_{ij} + \lambda other\,controls_{ij} + \varepsilon_{ij}, \end{aligned} \tag{1}$$

where infant mortality is an individual level mortality indicator for live birth i in primary sampling unit j , taking values 0 if the child survived the first 12 months of life and 1000 if the child died (this helps rescale it to the convention of reporting infant mortality rates per 1,000). We are primarily interested in the coefficient β_1, a binary indicator for whether the child lives in a rural area, and the extent to which measures of socioeconomic status, environmental exposures, maternal characteristics, and access to healthcare can reduce the magnitude and significance of the coefficient. Throughout, we cluster standard errors at the level of the primary sampling unit, and adjust for survey weights. All continuous variables are standardized, such that their effects should be interpreted as the effect of one standard deviation change. Because some of the explanatory variables may be correlated with each other, we introduce socioeconomic status or environmental exposures one at a time. In fuller models, we examine if the environmental exposures could still be significant after controlling for measures of socioeconomic status, which is a test of their predictive ability for the outcome in comparison to socioeconomic status variables.

Next, we use a two-way Blinder (1973)-Oaxaca (1973) decomposition, as implemented in Stata by Jann et al. (2008), to estimate the portion of the difference in infant mortality in rural and urban India that can be explained by the differences in the magnitudes of the explanatory variables. The unexplained difference is due to the effects of the explanatory variables and their interactions.

$$infant\,mortality_{ij} = \begin{cases} \beta^{rural} Explanatory\,factors_{ij} + \varepsilon_{ij}^{rural} & if rural \\ \beta^{urban} Explanatory\,factors_{ij} + \varepsilon_{ij}^{urban} & if urban \end{cases} \tag{2}$$

$$(0.5\beta^{rural} + 0.5\beta^{urban})(\overline{Explanatory\,factors_{ij}^{rural}} - \overline{Explanatory\,factors_{ij}^{urban}}) \tag{3}$$

3 Results

3.1 Descriptive graphs

Figure 1 shows results from local polynomial regressions for rural and urban infant mortality rates by some of the exposures and measures of socioeconomic status, allowing us to examine these relationships non-parametrically. Sub-figure 1 shows that higher maternal education is associated with lower infant mortality, and at all levels of maternal education, rural areas have a disadvantage. With maternal age, the figure documents a u-shaped relationship, with the u-shape being stronger in rural areas. Given this non-linear relationship, the regressions adjust for maternal age by including dummies for 5 year age intervals. Still, at all levels of maternal age, rural areas have a disadvantage.

Both household and neighbourhood level asset wealth is associated with steeper mortality declines in rural areas, and at very high levels of asset wealth, respondents in rural areas are not worse off than respondents of urban areas. Interestingly, household asset wealth is not associated with declines in infant mortality for large portions of the asset distribution in urban areas, but neighborhood wealth is. This pattern is different than that of maternal education, and surprising in-itself, given prior research (Marmot, 2005).

Figure 2 considers environmental exposures linked to open defecation, and household and ambient air pollution. As expected, in both urban and rural areas, worse neighborhood sanitation is associated with higher child mortality, but few areas in urban areas have higher open defecation than .5, reflected in the large 95% confidence intervals. In rural areas, higher open defecation is linearly associated with higher infant mortality.

Unlike previous graphs, graphs that show rural-urban differences by neighborhood solid fuel use document that accounting for solid fuel use can explain a large portion of rural disadvantage. Indeed, by neighborhood solid fuel use, in areas where less than half of the neighbors use solid fuels, urban and rural areas are no different. Similarly, among those who use clean fuels, neighborhood solid fuel use can account for almost all the rural-urban difference. Among those who use solid fuels, there is still a difference at the higher end of the neighborhood solid fuel use distribution.

3.2 Regression results

Table 2 examines the rural infant mortality disadvantage in a parametric regression setting. All models control for birth order, mother's age in single years, month of birth, and year of birth as dummies. While the summary statisics showed that rural areas had 21 deaths more than urban areas per 1,000 live births, after accounting for demographic variables and sex of the child, rural areas had 18.6 more deaths per 1,000. Model 2 adjusts for household and neighborhood solid fuel use, and shows that this large rural-urban gap becomes close to zero and non-significant when solid fuel use is accounted for. Subsequent models show that no other explanatory factor can explain this big a gap in rural-urban areas, and that household and neighourhood open defecation reduces the gap most substantially after solid fuel exposure (model 6). Accounting for household and neighborhood open defecation (model

6) also reduces the urban advantage more than accounting for household asset index or maternal education. Model 3 adjusts for household wealth index, model 4 for access to health care, and model 5 for maternal education. Adjusting for maternal education in model 4 reduces the rural penalty to 13 deaths per 1,000 live births.

In Table 3, we examine the extent to which solid fuel use is a proxy for socioeconomic status or just women's status. Model 1 adjusts for household and neighborhood solid fuel use, and household and neighborhood wealth index. Notice that in Model 1 household solid fuel use does have a direct impact on infant mortality once wealth is controlled for. This implies that while household solid fuel use was a proxy for wealth, neighourhood solid fuel wasn't. Model 2 controls for health access as well, Model 3 for household and neighborhood open defecation (again, as expected, household open defecation is not associated with infant mortality, while neighourhood open defecation is), model 4 for mother's education and body mass index (as proxies for women's status as well), and model 5 for household and neighourhood electrification (which are not associated with infant mortality). Notice that the neighourhood solid fuel use coefficient remains stable around 5 per 1,000, and remains singificant as well.

3.3 Oaxaca-Blinder decomposition results

The Oaxaca-Blinder regression results reveal a similar story. Differences in the distribution of solid fuel use in rural and urban areas explains 16.76 deaths per 1,000, or about 80.2% of the total difference of 20.9 deaths per 1,000. Other explanatory variables can at best explain only about half the difference as explained by solid fuel use, and the addition of these variables to solid fuel use in the decomposition models does not increase the total difference explained. Additionally, household open defecation, household solid fuel use, and neighourhood wealth are not significantly related to differences in rural and urban areas.

4 Further work

These findings will be strengthened by additional analyses. Although pollution measures cannot be merged with earlier DHS surveys because of a lack of geographical identifiers in earlier surveys, the extent to which the environmental exposures can explain urban-rural infant mortality differentials can be examined from earlier DHS surveys (in 1992-93, 1998-99, 2005-06). It is possible that some of the explanatory variables, such as health-care coverage, may account for a greater proportion of the rural-urban disadvantage. This research will benefit by integrating monthly pollution datasets, which may allow the examination of seasonal impacts of pollution. It is also possible to examine neo-natal and post-neonatal mortality separately from the DHS survey, which I intend to do in subsequent versions. As of now, we have only considered urban-rural definitions as provided by the DHS, it might be useful to consider alternative ways of classifying areas into rural and urban (Wolfe and Behrman, 1982). The local polynomial regressions point to interactions between the exposures, rural-urban residence, and measures of socio-economic status. In particular, it seems that the exposures and measures of socioeconomic status drive the relationship more steeply in rural areas. In further work, I plan to estimate these interactions. Similarly, I intend to stratify regression equation 1 with rural status. I also intend to use nonparametric decomposition methods to decompose the difference between urban and rural areas as explained by the covariates and the exposures. I also intend to integrate other explanatory variables, such as water sources (Bhalotra et al., 2017), vaccination coverage (McGovern and Canning, 2015), pre-natal care, father's education Behrman and Rosenzweig (2002), and other measures of maternal health. In future analyses, I will keep only one birth per women. Since births only for births in the last two years are analysed, the number of such births is likely to be small. I will also estimate the regressions separately for the outcomes of neonatal and post-neonatal motality, to understand disease pathways.

5 Discussion

This paper investigates the extent to which environmental exposures, maternal characteristics, and measures of socioeconomic status can explain rural disadvantage in developing countries. It finds that just one environmental exposure - solid fuel use for cooking - can explain urban large portions of rural infant mortality disparities when its externalities are taken into account. This finding has important theoretical and policy implications. Theoretically, this paper documents that even though unlike historical developed country contexts where urban areas had a mortality penalty, rural areas in contemporary developing countries have a mortality penalty, the reasons for the mortality penalty in rural areas of India may be the same as the mortality penalty of urban areas in historical Europe and North America. Indeed, environmental exposures in India may be even more important than they were in historical Europe - Beach and Hanlon (2016) finds that coal use in Britain could explain one-third of the total urban penalty in 1851 - 1860. From a policy perspective, this paper offers evidence of the large costs of solid fuel use for child health in rural areas. While improving rural health-care provision receives substantial attention in rural areas, the conversation on addressing pollution is restricted to urban areas. This is partly because the only real-time pollution monitors are in urban areas, and urban residents have a greater ability to recognize and communicate harms of air pollution. Air pollution in rural areas is, however, both higher, and a substantial contributor to rural disadvantage. By separating the standard DHS wealth index into household wealth measures and environmental exposures, this research has also tried to uncover mechanisms through which socioeconomic status affects child health.

References

Adler, Nancy E, W Thomas Boyce, Margaret A Chesney, Susan Folkman, and S Leonard Syme. 1993. "Socioeconomic inequalities in health: no easy solution." *Jama*, 269(24): 3140–3145.

Bassani, Diego G, Prabhat Jha, Neeraj Dhingra, and Rajesh Kumar. 2010. "Child mortality from solid-fuel use in India: a nationally-representative case-control study." *BMC public health*, 10(1): 491.

Beach, Brian and W Walker Hanlon. 2016. "Coal smoke and mortality in an early industrial economy." *The Economic Journal*.

Behrman, Jere R and Mark R Rosenzweig. 2002. "Does increasing women's schooling raise the schooling of the next generation?" *American Economic Review*, 92(1): 323–334.

Bhalotra, Sonia R, Alberto Diaz-Cayeros, Grant Miller, Alfonso Miranda, and Atheendar S Venkataramani. 2017. "Urban water disinfection and mortality decline in developing countries."Technical report, National Bureau of Economic Research.

Blinder, Alan S. 1973. "Wage discrimination: reduced form and structural estimates." *Journal of Human resources*: 436–455.

Bocquier, Philippe, Nyovani Janet Madise, and Eliya Msiyaphazi Zulu. 2011. "Is there an urban advantage in child survival in sub-Saharan Africa? Evidence from 18 countries in the 1990s." *Demography*, 48(2): 531–558.

Caldwell, John and Peter McDonald. 1982. "Influence of maternal education on infant and child mortality: levels and causes." *Health policy and education*, 2(3): 251–267.

Chafe, Zoë A, Michael Brauer, Zbigniew Klimont, Rita Van Dingenen, Sumi Mehta, Shilpa Rao, Keywan Riahi, Frank Dentener, and Kirk R Smith. 2014. "Household cooking with solid fuels contributes to ambient PM2. 5 air pollution and the burden of disease." *Environmental health perspectives*, 122(12): 1314.

Chen, Shaohua and Martin Ravallion. 2007. "Absolute poverty measures for the developing world, 1981–2004." *Proceedings of the National Academy of Sciences*, 104(43): 16757–16762.

Condran, Gretchen A and Eileen Crimmins. 1980. "Mortality differentials between rural and urban areas of states in the northeastern United States 1890–1900." *Journal of Historical Geography*, 6(2): 179–202.

Cutler, David and Grant Miller. 2005. "The role of public health improvements in health advances: the twentieth-century United States." *Demography*, 42(1): 1–22.

Elo, Irma T. 2009. "Social class differentials in health and mortality: Patterns and explanations in comparative perspective." *Annual Review of Sociology*, 35: 553–572.

Geruso, Michael and Dean Spears. 2018. "Neighborhood sanitation and infant mortality." *American Economic Journal: Applied Economics*, 10(2): 125–62.

Government of India. 2011. "Census of India 2011." *Office of the Registrar General Census Commissioner, India. Indian Census Bureau.*

Haines, Michael R. 2001. "The urban mortality transition in the United States, 1800-1940." in *Annales de démographie historique*(1): 33–64, Belin.

Hippocrates. 2004. *On Airs, Waters and Places*: MIT Classics Library.

Hubbard, William H. 2000. "The urban penalty: towns and mortality in nineteenth-century Norway." *Continuity and Change*, 15(2): 331–350.

IIPS & ICF. 2017. "NFHS-4 (National Family Health Survey-4)." *International Institute for Population Studies.*

Jann, Ben et al.. 2008. "The Blinder-Oaxaca decomposition for linear regression models." *The Stata Journal*, 8(4): 453–479.

Kesztenbaum, Lionel and Jean-Laurent Rosenthal. 2011. "The health cost of living in a city: The case of France at the end of the 19th century." *Explorations in Economic History*, 48(2): 207–225.

Macintyre, Sally, Anne Ellaway, and Steven Cummins. 2002. "Place effects on health: how can we conceptualise, operationalise and measure them?" *Social science & medicine*, 55(1): 125–139.

Marmot, Michael. 2005. "Social determinants of health inequalities." *The lancet*, 365(9464): 1099–1104.

McGovern, Mark E and David Canning. 2015. "Vaccination and all-cause child mortality from 1985 to 2011: global evidence from the Demographic and Health Surveys."

American journal of epidemiology, 182(9): 791–798.

Mosley, W Henry and Lincoln C Chen. 1984. “An analytical framework for the study of child survival in developing countries.” *Population and development review*, 10: 25–45.

Oaxaca, Ronald. 1973. “Male-female wage differentials in urban labor markets.” *International economic review*: 693–709.

Pörtner, Claus C and Yu-hsuan Su. 2018. “Differences in child health across rural, urban, and slum areas: Evidence from India.” *Demography*, 55(1): 223–247.

Preston, Samuel H and Michael R Haines. 1991. *Fatal years: Child mortality in late nineteenth-century America*: Princeton University Press.

Reher, David S. 2001. “In search of the urban penalty: exploring urban and rural mortality patterns in Spain during the demographic transition.” *Population, Space and Place*, 7(2): 105–127.

Rehfuess, Eva, Sumi Mehta, and Annette Prüss-Üstün. 2006. “Assessing household solid fuel use: multiple implications for the Millennium Development Goals.” *Environmental health perspectives*, 114(3): 373.

Saikia, Nandita, Abhishek Singh, Domantas Jasilionis, and Faujdar Ram. 2013. “Explaining the rural-urban gap in infant mortality in India.” *Demographic Research*, 29: 473–506.

Sastry, Narayan. 1997. “What explains rural-urban differentials in child mortality in Brazil?” *Social science & medicine*, 44(7): 989–1002.

Smith, Kirk R, John P McCracken, Martin W Weber, Alan Hubbard, Alisa Jenny, Lisa M Thompson, John Balmes, Anaité Diaz, Byron Arana, and Nigel Bruce. 2011. “Effect of reduction in household air pollution on childhood pneumonia in Guatemala (RESPIRE): a randomised controlled trial.” *The Lancet*, 378(9804): 1717–1726.

Snow, John. 1856. “On the mode of communication of cholera.” *Edinburgh medical journal*, 1(7): 668.

Van de Poel, Ellen, Owen ODonnell, and Eddy Van Doorslaer. 2007. "Are urban children really healthier? Evidence from 47 developing countries." *Social science & medicine*, 65(10): 1986–2003.

Van de Poel, Ellen, Owen O'donnell, and Eddy Van Doorslaer. 2009. "What explains the rural-urban gap in infant mortality: household or community characteristics?" *Demography*, 46(4): 827–850.

Van Donkelaar, Aaron, Randall V Martin, Michael Brauer, N Christina Hsu, Ralph A Kahn, Robert C Levy, Alexei Lyapustin, Andrew M Sayer, and David M Winker. 2016. "Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors." *Environmental science & technology*, 50(7): 3762–3772.

Wolfe, Barbara L and Jere R Behrman. 1982. "Determinants of child mortality, health, and nutrition in a developing country." *Journal of Development Economics*, 11(2): 163–193.

World Health Organization. 2015. *Progress on sanitation and drinking water: 2015 update and MDG assessment*: World Health Organization and WHO/UNICEF Joint Water Supply and Sanitation Monitoring Programme.

You, Danzhen, Lucia Hug, Simon Ejdemyr, and Jan Beise. 2017. "Levels and trends in child mortality. Report 2017. Estimates developed by the UN Inter-agency Group for Child Mortality Estimation.."

Figure 1: Local polynomial regressions examining association of maternal characteristics, socioeconomic status, neighborhood sanitation with infant mortality in urban and rural areas

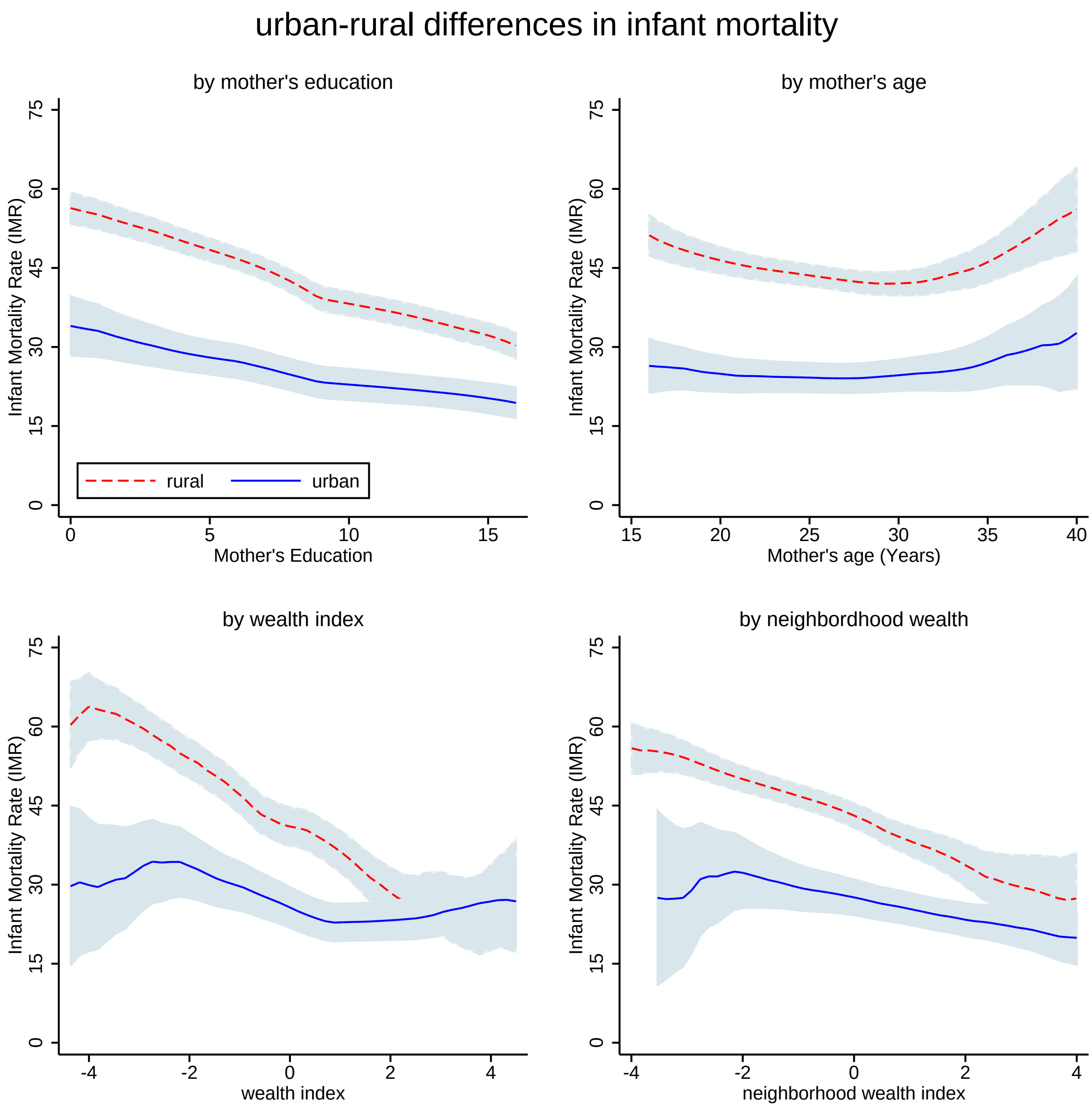

Figure 2: Local polynomial regressions examining association of satellite derived air pollution levels and neighborhood solid fuel use with infant mortality in urban and rural areas

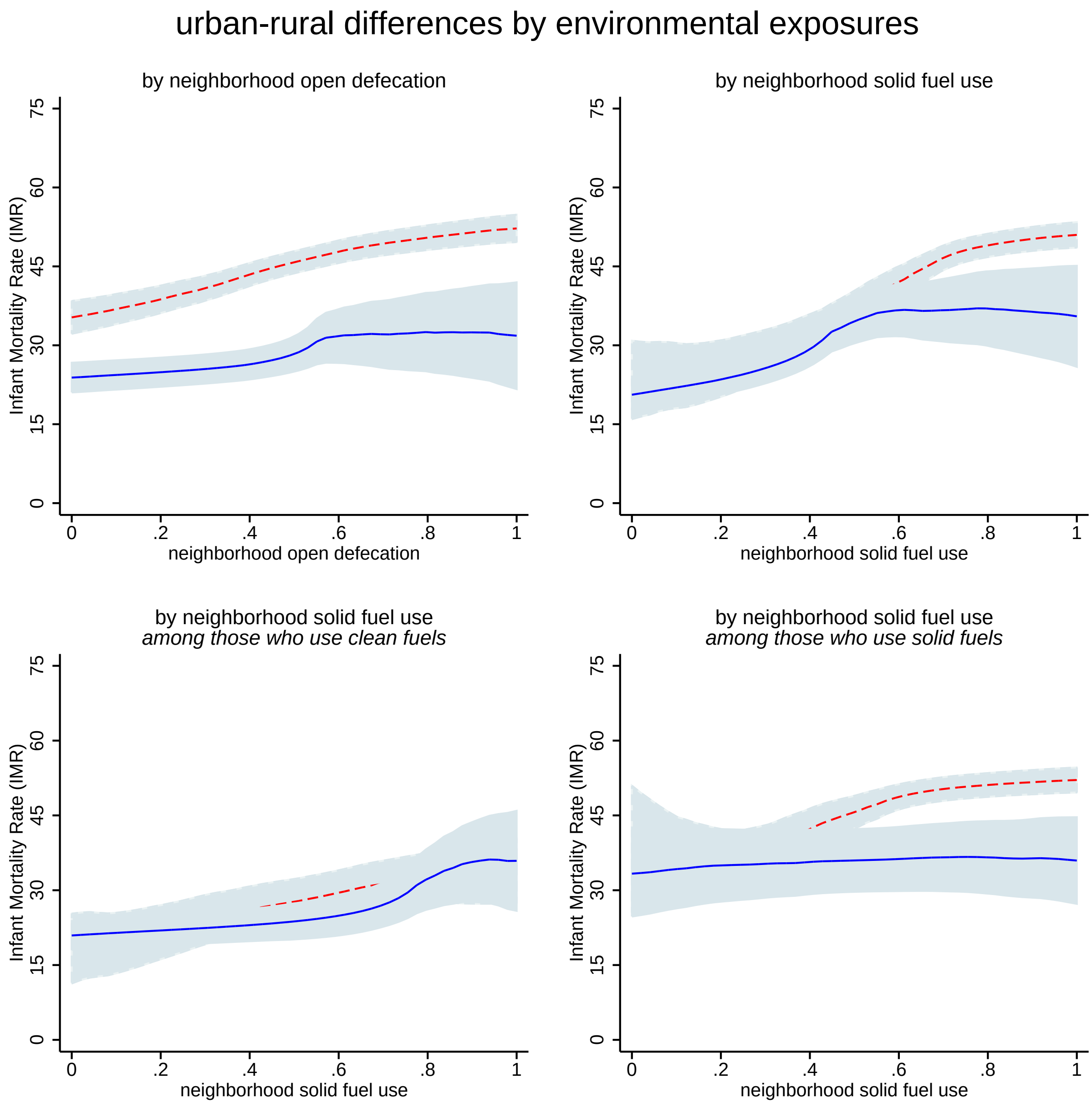

Table 1: Summary statistics for the estimation sample

	Total		Urban		Rural	
	mean	s.e.	mean	s.e.	mean	s.e.
outcome						
IMR	40.3	1.11	25.2	1.77	46.1	1.37
demographic covariates						
female	0.48	0.00	0.48	0.01	0.48	0.00
birth order	2.2	0.01	1.9	0.01	2.3	0.01
maternal characteristics						
mother's age	25.5	0.02	25.9	0.06	25.3	0.02
mother's bmi	20.9	0.02	22.2	0.05	20.4	0.02
mother's education	6.7	0.03	8.8	0.07	5.9	0.03
measures of socioeconomic status						
wealth index	-0.17	0.01	1.23	0.03	-0.68	0.01
meanwealth index in psu	-0.37	0.01	1.15	0.03	-0.93	0.01
electrification	0.84	0.00	0.97	0.00	0.80	0.00
neighourhood electrification	0.83	0.00	0.97	0.00	0.78	0.00
health access at time of delivery						
delivery by doctor / nurse	0.70	0.00	0.80	0.01	0.67	0.00
delivery by health worker at facility	0.13	0.00	0.12	0.01	0.14	0.00
delivery by health worker at home	0.15	0.00	0.08	0.01	0.18	0.00
home birth (without health worker)	0.01	0.00	0.01	0.00	0.01	0.00
environmental exposures						
open defecation	0.47	0.00	0.14	0.01	0.59	0.00
neighourhood open defecation	0.47	0.00	0.14	0.00	0.60	0.00
mean pm 2.5 levels	50.4	0.14	47.6	0.34	51.4	0.15
solid fuel use	0.66	0.00	0.24	0.01	0.81	0.00
Neighourhood solid fuel use	0.64	0.00	0.22	0.01	0.80	0.00
n (births)			54,722			

Note: observations are births between 12 and 24 months before the interview date. Standard Errors and means reflect the complex survey naturs of the DHS, and are calculated using STATA's svy routine, accounting for clustering, strata, and survey weights. psu = primary sampling unit.

Table 2: OLS regressions examining the extent to which exposures and covariates can explain rural infant mortality disadvantage

	(1)	(2)	(3)	(4)	(5)	(6)
rural	18.57***	0.645	11.67***	16.58***	13.51***	7.707**
	(2.257)	(3.262)	(2.351)	(2.281)	(2.301)	(2.657)
female	-7.030***	-7.246***	-7.381***	-7.187***	-7.156***	-7.142***
	(2.111)	(2.110)	(2.113)	(2.109)	(2.109)	(2.108)
household solid fuel use		10.48***				
		(3.082)				
neighborhood solid fuel use		7.229***				
		(1.812)				
household wealth index			-9.898***			
			(1.164)			
health access at delivery *(Ref: Doctor or Nurse in health facility)*						
health worker (at facility)				3.519		
				(3.382)		
health worker (at home)				20.42***		
				(3.681)		
home birth (without health worker)				80.48***		
				(17.46)		
mother's education					-11.33***	
					(1.234)	
household open defecation						9.489**
						(3.547)
neighborhood open defecation						5.873**
						(1.903)
constant	23.33***	29.78***	26.49***	18.90***	23.63***	24.99***
	(5.329)	(6.008)	(5.345)	(5.393)	(5.321)	(5.614)
mean of dependent variable	40.3	40.3	40.3	40.3	40.3	40.3
se of dependent variable	(1.11)	(1.11)	(1.11)	(1.11)	(1.11)	(1.11)
N	54722	54722	54722	54722	54722	54722

Note: Clustered standard errors, clustered at the psu level, in parantheses. + p<0.1, * p<0.05, ** p<0.01, *** p<0.001. All estimates are weighted using survey weights. Source: DHS India 2014-15. All models adjust for mother's age in single years, month of birth, year of birth and birth order as dummies.

Table 3: OLS regressions examining robustness of role of solid fuel use exposure in explaining rural infant mortality disadvantage

	(1)	(2)	(3)	(4)	(5)
rural	1.004	2.379	0.776	1.752	2.079
	(3.252)	(3.231)	(3.261)	(3.279)	(3.275)
household solid fuel use	4.608	4.175	3.652	3.784	3.785
	(3.247)	(3.237)	(3.250)	(3.268)	(3.269)
neighborhood solid fuel use	7.138**	6.347**	5.155*	4.961*	4.898*
	(2.216)	(2.198)	(2.174)	(2.174)	(2.174)
household wealth index	-8.318***	-7.537***	-7.050***	-5.340**	-4.651**
	(1.492)	(1.486)	(1.566)	(1.712)	(1.764)
neighborhood wealth index	2.393	3.093+	5.726**	5.756**	7.093***
	(1.880)	(1.870)	(2.089)	(2.080)	(2.153)
health access at delivery *(Ref: Doctor or Nurse in health facility)*					
health worker (at facility)		2.281	1.673	2.211	1.830
		(3.371)	(3.376)	(3.375)	(3.373)
health worker (at home)		15.19***	15.36***	14.37***	13.13***
		(3.712)	(3.708)	(3.701)	(3.719)
home birth (without health worker)		75.91***	75.88***	74.22***	72.81***
		(17.56)	(17.56)	(17.49)	(17.60)
household open defecation			2.877	2.523	2.433
			(3.801)	(3.799)	(3.813)
neighborhood open defecation			14.08*	15.23*	13.79*
			(6.010)	(6.012)	(6.022)
mother's education				-7.407***	-7.313***
				(1.472)	(1.474)
mother's body mass index				9.238***	9.191***
				(1.494)	(1.495)
square of mother's body mass index				-1.869***	-1.863***
				(0.533)	(0.533)
household electrification					-5.972
					(4.220)
neighborhood electrification					-2.448
					(1.711)
constant	32.23***	27.10***	20.20**	19.68**	25.41***
	(6.019)	(6.071)	(6.380)	(6.398)	(7.171)
N	54722	54722	54722	54722	54722

Note: Clustered standard errors, clustered at the psu level, in parantheses. + $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All estimates are weighted using survey weights. Source: DHS India 2015-16. All models adjust for mother's age in single years, month of birth, year of birth and birth order as dummies. Coefficients for female not shown.

Table 4: Results from Oaxaca-Blinder decompositions

Overall Gap							
Rural	46.10***	Urban	25.20***	Difference	20.90***	To explain	20.90***
	(1.370)		(1.765)		(2.235)		(2.235)

The extent to which rural urban gap is explained by:							
Wealth		Maternal characteristics		Open defecation		Solid fuel use	
household	2.302	education	4.338***	household	7.763	household	2.134
	(1.737)		(1.159)		(4.973)		(3.228)
neighborhood	3.484	bmi	0.323	neighborhood	0.450	neighborhood	14.62**
	(2.963)		(0.762)		(6.182)		(4.939)
		age	-0.179				
			(0.223)				
Total explained							
5.787*		4.481***		8.213*		16.76***	
(2.356)		(1.162)		(3.396)		(3.697)	

The extent to which the rural urban gap is explained by covariates in addition to solid fuel use:							
Solid fuel use + wealth		Solid fuel use + maternal characteristics		Solid fuel use + open defecation		Solid fuel use + child characterisitcs	
household solid fuel use	0.982 (3.420)	household solid fuel use	0.373 (3.372)	household solid fuel use	0.296 (3.616)	household solid fuel use	1.359 (3.250)
neighborhood solid fuel use	16.70** (5.789)	neighborhood solid fuel use	12.81** (4.946)	neighborhood solid fuel use	18.10** (5.543)	neighborhood solid fuel use	13.80** (4.960)
household wealth	2.213 (1.841)	mother's education	3.343** (1.232)	household open defecation	7.676 (5.294)	birth order	1.263* (0.571)
neighborhood wealth	-3.132 (3.581)	mother's bmi	0.00883 (0.778)	neighborhood open defecation	-10.05 (6.418)	month of birth	0.0637 (0.0662)
		mother's age	-0.202 (0.223)			year of birth	-0.0145 (0.0332)
Total explained							
16.77***		16.33***		16.02***		16.47***	
(3.698)		(3.685)		(4.055)		(3.725)	

Note: Clustered standard errors, clustered at the psu level, in parantheses. + p<0.1, * p<0.05, ** p<0.01, *** p<0.001. All estimates are weighted using survey weights. Source: DHS India 2014-15.

Figure A1: Satellite derived PM 2.5 levels, 2014

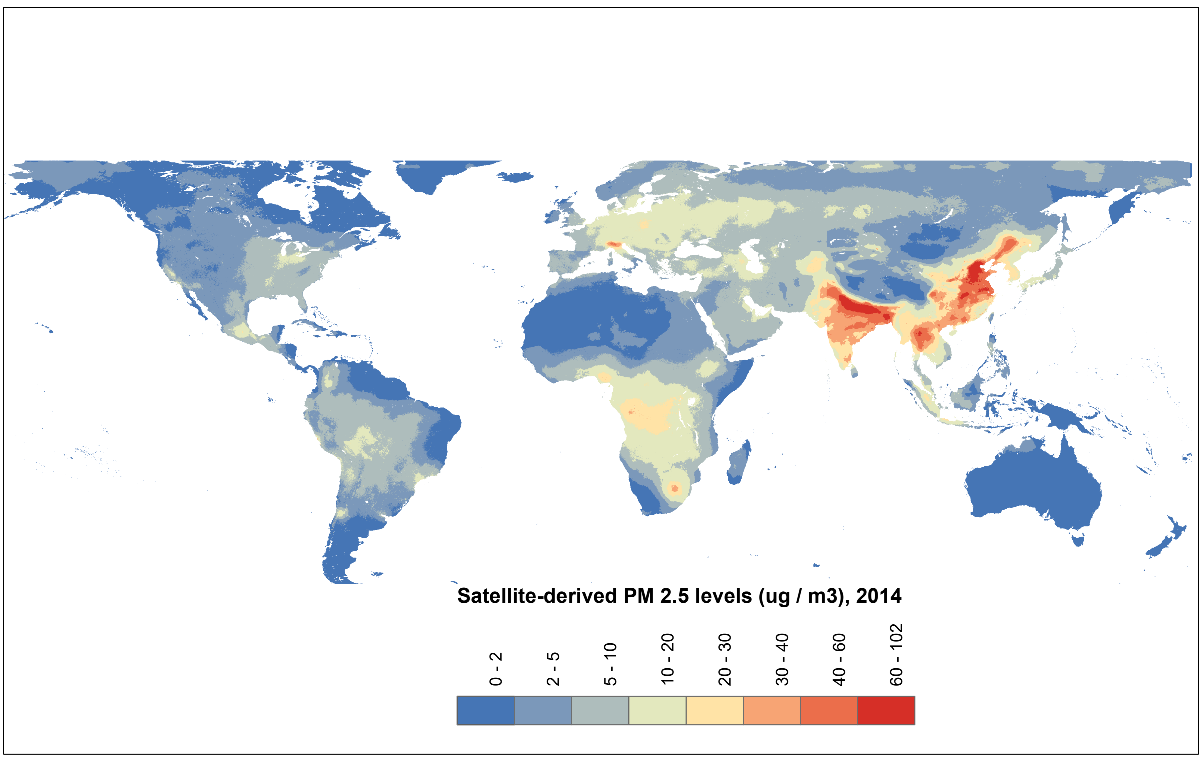

Figure A2: DHS 2015-16 clusters (green dots) mapped to PM 2.5 levels, 2014

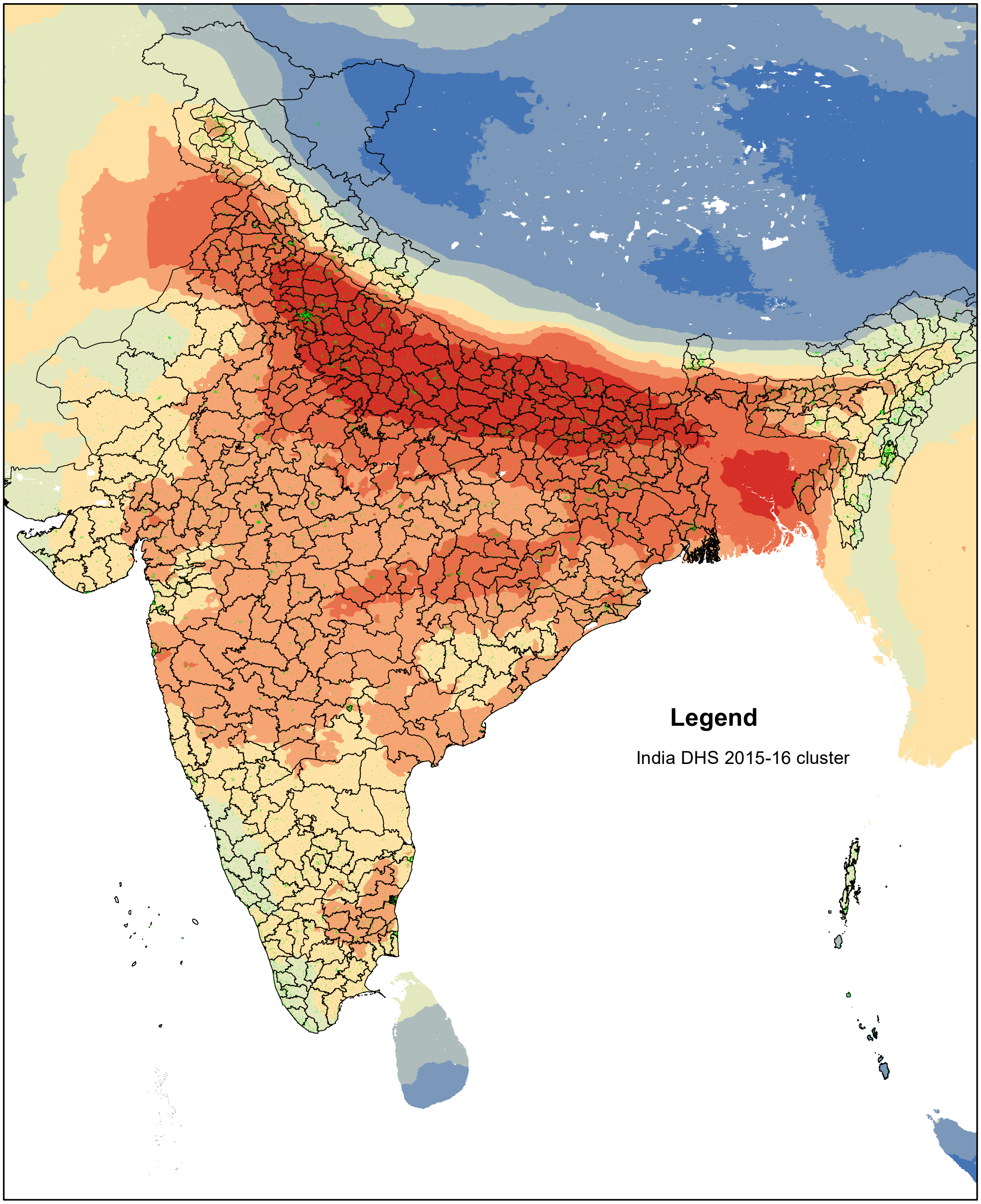

Figure A3: neighborhood solid fuel use predicts air pollution in rural areas