The Role of Forests in Reducing the Incidence of Acute Respiratory Infections: Evidence from India

Andrew Miller* Daniela Miteva Sathya Gopalakrishnan

1. Introduction

Manifesting with symptoms such as fever, cough, shortness of breath, wheezing and difficulty breathing, Acute Respiratory Infections (ARI) are an important health risk that has been estimated to account for as much as 7.0% of total deaths in low and middle-income countries (Lopez et al., 2006). While they most commonly caused by viruses or mixed viral-bacterial infections, (World Health Organization, 2014), previous work has highlighted the role of air quality, both indoors and outdoors, as a dominant risk factor (e.g. Cortes-Ramirez et al. 2021). For instance, because of the increase in particulate matter and exacerbated outdoor air quality, forest fires in Indonesia in the mid-1990's have been estimated to have caused a half-standard deviation decrease in the height-for-age z score at age 17, implying a 4% loss of average monthly wages for the 1 million Indonesian workers born during this period (Tan-Soo and Pattanayak 2019).

Forests have been linked to a host of health benefits in the global health literature, including the reduction of ARI. Three main mechanisms relate forests to the incidence of ARI. First, trees can directly remove particulate matter, specifically $PM_{2.5}$ and PM_{10}, from the air (Nowak et al., 2002, 2014; Escobedo et al., 2008; Ratajczak et al., 2021). $PM_{2.5}$ and PM_{10} are measurements of the density of particulate matter present in the air of less than 2.5 micrometers and 10

*Corresponding author: miller.5776@osu.edu

micrometers in diameter respectively[1]. This particulate matter may enter and irritate the lungs, with $PM_{2.5}$ being particularly harmful due to its small diameter size and resultant ability to enter the lungs and circulatory system, causing inflammation, decreased lung function and increased susceptibility to infections such as those that comprise ARI (Larson et al., 2022).

Second, contact with forests may improve the human immune system directly. Time spent in forest environments has been shown to alleviate stress and anxiety, which in turn lowers cortisol levels and promotes proper functioning of the immune system (Lee et al., 2011; Wen et al., 2019). While relatively nascent, there is growing evidence that the biogenic volatile organic compounds (BVOCs) emitted by forest trees may themselves have therapeutic effects in reducing respiratory inflammation and improve respiratory health (Kim et al., 2020), though this effect can be heterogenous, as short-term exposure to certain BVOCs can themselves have negative effects on human health, and urban BVOCs can exacerbate the effects of Anthropogenic Volatile Organic Compounds (Lun et al., 2020).

Third, in areas with degraded forests that still have vegetation, people may resort to using lower quality fuelwood, exacerbating the negative effects of the use of fuelwood for cooking and heat (Jagger and Shively, 2014).

While theory suggests forests can help mitigate some of the ARI disease burden, the empirical evidence of the magnitude of the impact is still very scarce and appears context specific. For example, Bauch et al. (2015) analyze municipal level data from the Brazilian Amazon using a random effects quantile regression. They find that ARI among children under the age of two is significantly and negatively correlated with string environmental protection of forests. Working with data from the Cambodian Demographic and Health Surveys (DHS), Pienkowski et al. (2017) use a generalized linear mixed effects model to estimate the the relationship between the deforestation of dense forest and ARI incidence in children, finding a positive and highly statistically significant relationship. Pattanayak et al. (2010) use original household survey data from Odisha, India to estimate a negative relationship between the diversity of the forest within two kilometers of the village and days of illness. Surprisingly, they also find

[1]For reference, human hair has a diameter of 50-70 micrometers.

a positive relationship between forest stock within two kilometers of the village and number of household days ill, a result which they attribute to their variable not capturing the full benefits of forests since forests outside of a two kilometer radius may convey substantial benefits. In the study most similar to ours in approach and data, Berazneva and Byker (2017) use Nigerian DHS survey data and remotely sensed forest cover from Hansen et al. (2013), to estimate the relationship between forest loss and ARI symptoms in children under five. They find no statistically significant relationship.

Here we use data from the spatially explicit Demographic and Health Surveys (DHS) from the 2019 wave in India (The DHS Program, 2019-21) to empirically test the relationship between forest cover and the ARI incidence in children under five. In order to account for the hierarchical structure of our data, we test this relationship using a hierarchical linear model, with levels at the district, survey cluster and household level. In our preferred specification, we find a small but statistically significant and negative relationship between forest cover and ARI symptoms. That is, for an additional square kilometer of forest cover at the district level, the share of rural children experiencing ARI symptoms is reduced by -0.00322. While size of this effect at the individual level is small, it translates to significant health impacts when aggregated to the district and national levels. Putting this into monetary terms, we estimate an average value per square kilometer of $8,922, with a total value of $27,555,051 across the districts in our sample. Our work builds upon prior work in three important aspects. First, we focus on India, which has a high burden of ARI infections. In our sample of rural children under the age of five, 6.5% of children experienced symptoms of ARI in the two weeks leading up to the survey. India has also lost 5.3% of its forest cover between 2000 and 2021 (Global Forest Watch). Estimation of the relationship between forests and ARI in the specific context of India is essential of public officials are to make fully informed decisions regarding the benefits associated with the protection of forested areas while balancing the needs of a developing economy. Second, by using hierarchical linear model (HLM) estimation to account for unobserved variables, the estimates that we provide are at least partially identified, giving policymakers increased confidence in their value. Finally, we estimate the monetary value of the estimated reduction in ARI symptoms, allowing our estimates to be plugged in to ongoing ecosystem service valuation efforts for India's forests.

2. Methods

We control for household and village-level covariates that may affect respiratory health, especially those related to indoor air quality. The use of fuelwood for indoor cooking is an important factor in determining indoor air quality, especially among children who are often supervised by mothers cooking for their households (Pattanayak and Pfaff, 2009). Ezzati and Kammen (2001) estimate that indoor smoke from solid fuels is responsible for 2.7% of the global disease burden, and Lewis et al. (2017) estimate that use of improved cook stoves are associated with >70% reduction in indoor pollutants, reduced payments for treatment of ARI, and reduced time spent in the hospital for ARI. Another important contributor to poor indoor air quality is the presence of smoking within the household Merera (2021); Varghese and Muhammad (2023), which we control for in our estimation. Younger children are typically more susceptible to ARI Merera (2021), and we include the mean age of the surveyed children within the household. For socioeconomic variables, we include the mother's education, and whether or not the household belongs to a scheduled caste or tribe. We have excluded household wealth due to issues of near-multicollinearity with the included household variables known to be causal predictors of ARIs, which can be a particularly complex problem when using hierarchical models (Yu et al., 2015). At the village/cluster level, we included the altitude of the village since altitude may be a determinant in ARI risk (Hasan et al., 2022).

We are interested in the impact of forests on the incidence of ARIs and expect the forest variable to reduce the incidence of ARIs. We calculate the forest area at the district level, and use the log-transformed number of square kilometers. Because the unit of analysis is an individual household, it is unlikely that there is reverse causality between ARI prevalence and the forest variable, but we do account for potentially bias due to omitted variables via control function estimation (Wooldridge, 2015).

2.1 Hierarchical linear model

In order to account for the of the nested structure of our data, we estimate the effect of deforestation on ARI symptoms using a hierarchical linear model (HLM) (Raudenbush and Bryk, 2002). This allows us to account for unobserved factors that are correlated with our outcome at multiple group levels and cause biased estimates when using the naive ordinary least squares estimation. In our case, we include a random intercept at the district and DHS cluster (analogous to village) levels.

With the inclusion of random intercepts at the district, cluster and household levels, we have the following hierarchical linear model:

$$\begin{aligned} y_{hcd} &= \alpha_{0cd} + \beta X_{hcd} + \varepsilon_{hcd} \\ \alpha_{0cd} &= \alpha_{0cd} + \lambda Z_c + u_{0cd} \\ \alpha_{0d} &= \zeta_{00} + \psi U_d + u_{0d} \end{aligned} \tag{1}$$

where h indexes the household, c indexes the cluster and d indexes the district. X_{hcd} is the matrix of household control variables (mean child age, mother's education, clean cooking fuel, smoking in household, household belonging to a scheduled caste or tribe), Z_c is our matrix of cluster-level variables (altitude, month of survey interview), and U_d is the matrix of district-level variables, including our variable of interest, forest area. Essentially, at each level of the model, the expected proportion of a given household displaying ARI symptoms is adjusted away from the group proportion conditional on the characteristics of the group characteristics at that level, plus a remaining error term at the level. Thus, ζ_{00} is the grand proportion, and the random intercept of group d is determined by its deviation from that mean, partially explained by the matrix of group characteristics U_d. This is repeated for the other levels so that α_{0cd} is the cluster-level proportion for a child in cluster c, which is in district d.

To estimate equation 1, we used the `xtmixed` command from (STATA, 2019) with the inclusion of cluster-level fixed effects. Since the our forest cover data is calculated at the district level, we cluster our standard errors at the district level (Abadie et al., 2023). As discussed in detail in section 4, we do not include sample weights in the estimation of the HLM based on the suggestions of Solon et al. (2015) and our understanding of the nature survey-design induced endogenity. Specifically, we argue that our estimation of random intercepts controls for unobserved spatial heterogeneity.

2.2 Control function approach to address potential endogeneity

While our HLM estimates are plausibly causal, it is difficult to rule out omitted variable bias due to the cross-sectional nature of our data. While the DHS is conducted roughly every five years, there is no way reliable way to link clusters between series of data collection, much less households or individuals, rendering panel data econometric techniques unusable. In order to test and control for potential endogeneity in the above specification, we consider a control function approach (Wooldridge, 2015) using the number of square kilometers over which tigers have been observed in each district as an instrument for forest cover (Sanderson et al., 2023). The raw correlation between the presence of tigers and hectares of forest in a district is 0.369, and the Montiel Olea-Pflueger test suggests that our instrument is adequately strong (Olea and Pflueger, 2013), fulfilling the relevancy requirement. In terms of the exclusion restriction, we argue that it is highly unlikely that child respiratory illness would have an effect on the presence of tigers, particularly at the district level. In the first stage we have

$$\begin{aligned} Forest\ cover\ =&\beta_0 + \beta_1 Tiger\ area + \beta_2 Mother\text{'}s\ educ + \beta_3 Clean\ cooking \\ &+ \beta_4 Smoking + \beta_5 SC/ST + \beta_6 Female + \beta_7 Month + \beta_8 Altitude + \nu \end{aligned} \tag{2}$$

which we estimate to obtain $\hat{\nu}$. We can then estimate the second stage equation

$$\begin{aligned} ARI\ =&\delta_0 + \delta_1 Forest\ cover + \delta_2 Mother\text{'}s\ educ + \delta_3 Clean\ cooking \\ &+ \delta_4 Smoking + \delta_5 SC/ST + \delta_6 Female + \delta_7 Month + \delta_8 Altitude + \rho_1 \hat{\nu} + \varepsilon \end{aligned} \tag{3}$$

Here the null hypothesis $H_0 : \rho_1 = 0$ would indicate that there was no endogeneity to be controlled for. We use wild bootstrapped standard errors clustered at the district level to account for the fact that our variable of interest is measured at the district level (Roodman et al., 2019).

Finally, as a robustness check, we also estimated a negative binomial count model of the number of ARI-symptomatic children in the household in the two weeks leading up to the survey (Long, 2024). In addition to the covariates described above, we include the number of children in the household. Using similar control variables to our HLM model, we estimated a negative binomial model using STATA's `nbreg` command.

3. Study Area

While our study area is representative of all of rural India, our sample includes 286 distinct districts out of 707 present in India. Districts were excluded if they did not have rural observations in the year of 2019, or did not have at least 100 hectares of forest. The data are described further in section 4. The district-level incidence of ARI symptoms among children under five has significant variation within our sample, with district proportions ranging from 0% to 25.2% (see figure 1). The percentage of forest cover at the district level also varies widely, with some districts being dropped from the sample due to the almost complete lack of forest, and others exhibiting up to 92.3% forest cover, with a mean of 65,620.1 ha., and a standard deviation of 114,982.5 ha. Many of the districts in our sample saw significant deforestation in the year prior to the survey, while the overall district means were relatively low, with a district mean of 0.056% (sd=0.215%, see figure 6) in terms of share of district area. An important mechanism by which forest cover can improve respiratory health is by improving outdoor air quality, which we measure by estimated $PM_{2.5}$ concentrations from the Copernicus Atmospheric Monitoring Service (Inness et al., 2019). Figure 7 shows the number of days that each district exceeded the World Health Organization's (WHO) interim $PM_{2.5}$ air quality goal. The mean district value for this was 285.1342, with some districts not exceeding the target once, and others exceeding the benchmark every day of the year.

Though we are unaware of large public health interventions that would bias our estimates, there is a large network of development agencies working in India, and it unlikely that there were no efforts to improve respiratory health during the study period. However, any such efforts are likely to be channeled through the primary determinants of ARI, specifically the improvement of indoor air quality (via improved stove technology and smoking cessation), or improved access to treatment at the village level. Each of these variables are controlled for in our models.

4. Data

Our primary data come from the Indian Demographic and Health Survey (DHS), also known as the NFHS-5, conducted from June 2019 to April 2021 (DHS 2019-21). The DHS was a country-wide survey that contains individual observations on health outcomes of children ages five and under, as well as household level control variables. The data were collected via a multi-stage sampling design, with Phase-I occurring from June 2019 to January 2020 covering 17 states and 5 unincorporated territories. Phase-II occurred subsequently and covered the remaining 11 states and 3 unincorporated territories, but are not included in our sample due to concerns about the effect of COVID-19 on our outcome variable. Despite this, we can see in figure 1, that we have data for each region of the country. Only coordinates of the village ("cluster") from which households are selected, are provided in the data; to protect anonymity, the location of the rural clusters is randomly displaced by a distance of up to 5km; for 1% of the sample, the location is displaced up to 10km. However, the clusters are always located in their original district. Unfortunately, it is not possible to link households or clusters to those of previous DHS survey, so our analysis is cross-sectional. We focus on the rural households, and drop districts with less than 10 ha of forest cover. For these reasons, our final sample contains 47,712 observations in 9,513 rural clusters.

4.1 Survey weights

The DHS in India utilized stratified sampling to ensure adequate coverage of sub-populations to be able to estimate key statistics for those sub-populations at the district and state levels. A uniform sampling design was used across districts. Within districts, PSUs were stratified by urban/rural designation, village size, percentage of Scheduled Cast / Scheduled Tribe population and female literacy rate. In the rural areas, villages where used as Primary Sampling Units (PSUs) and were selected with probability proportional to size (PPS). In the second stage, lists of all occupied households within each PSU was used to randomly select an equal number of households within each PSU, with a final target sample of 20 households per PSU.

Whether it is better to use survey weights in the estimation of causal effects is dependent on the specifics of research question and sampling strategy. In some cases, the use of survey weights may be unnecessary if the control variables adequately capture all sources of endogeneity so that the expectation of the error term given the covariates is zero. In this case, the inclusion of survey weights may only serve to reduce the precision of the estimates. However, if there is endogeneity introduced by the sampling process that is not fully captured by the covariates, the unweighted coefficient estimates will be inconsistent. In this case, estimates weighted by the inverse sampling probabilty is needed to achieve consistent estimates, and these estimates should be accompanied by robust standard errors (Solon et al., 2015). In short, once heteroskedasticity is accounted for, the only disadvantage of weighting the estimates is a loss of precision. Bollen et al. (2016) provide a helpful review and recommendations of empirical tests that can be used to ascertain whether weighting is necessary. We conducted two of these tests, the first suggested by Pfeffermann and Sverchkov (2007) with results found in table 6, and the second due to Hausman (1978), with results found in table 7. The results of both of these tests, plus the recommendations of DHS staff suggest that the inclusion of survey weights is prudent in our control function estimation. However, in our estimation of the hierarchical linear model we do not include the survey weights due to the fact that the primary source of endogeneity introduced by the DHS sampling procedure is likely due to unobserved geographical and spatial characteristics, which the nested structure of the HLM estimation controls for.

4.2 Outcome

Our data contain observations of individual children under 5 years of age. We use the presence of ARI symptoms in the two weeks prior to the DHS survey interview. Specifically, the child must have had short, rapid breathing which was chest related or difficulty breathing which was chest related in the two weeks leading up to the survey. Figure 1 shows the proportion of the sample with ARI symptoms by district, where we note significant variation with percentages between 0.0% and 25.2%. From the binary indicator of the presence of ARI symptoms in the two weeks leading up to the survey, we calculate our primary outcome variable, the proportion of children within a household who have had ARI symptoms.

4.3 Socio-economic covariates

In order to control for individual and household level covariates which may be driving ARI incidence, we control for important factors. These include smoking, mother's education, household fuelwood or biomass consumption for cooking, and whether or not the household belongs to a scheduled caste or tribe. Due to high levels of multicollinearity between wealth, mother's education and cooking fuel, we chose to exclude the wealth variable.

4.4 Forest data

In order to quantify the forest area, we use the 30 by 30 m tree cover data from Hansen et al. (2013). We define forests as pixels that have at least 25% tree cover at the baseline (Sexton et al., 2016) and restrict any forest loss events to only cells considered forest at the baseline. To define the forest area in any given year, we subtract past forest loss events from the baseline layer. Note that this approach excludes any reforestation and afforestation events. We are not aware of other comparable datasets that span the whole country at the same resolution. We calculate the total forest area and the forest area lost in a given year at the district level. Since the timing of tree cover loss is not explicit within the year, we use forest loss in the full year preceding the survey interview as our forest loss variable of interest[2]. Figure 2 illustrates the forest cover in square kilometers across the study area. Figure 6 shows the amount of forest lost as a share of the district. Much of the variation in our forest loss data comes from northeastern India, with losses of up to 2.3% of district area.

5. Results

The estimated coefficients from the primary specification are reported in column 2 of table 1. While the size of the estimates are small in absolute terms, we see that there is a statistically significant effect (at the 1% level) in the hypothesized direction for our main variable of interest, logged square kilometers of forest in the district. This coefficient can be interpreted as the predicted change in the proportion of children with ARI symptoms in a surveyed household in the two weeks leading up to the survey due to an increase in the number of square kilometers of

[2]We calculated forest loss both in percentage terms of the district size and previously existing forest.

forest in the same district. Once aggregated, we argue that this effect has a significant impact on ARI prevalence at the district and national level. We note that inclusion in a Scheduled Tribe has a negative and statistically significant (at the 10% level) coefficient. One explanation for this is that there have been initiatives within India to institute community-managed forests in India where tribes have control over the use of forests (Miller et al., 2017), and there is some evidence that these community-managed forests are better conserved than their state-managed counterparts Somanathan et al. (2009). Within our sample, the district share of the households affiliated with a scheduled tribe was positively correlated with forest cover ($\rho = 0.27$).

To account for potential unobserved correlation between forest cover and the error term, we also estimated the model using the control function approach described in Wooldridge (2015). We see in table 2 the results of the second stage of this estimation. We see a very similar coeffecients to our HLM estimation. We also note that the first stage residual is not statistically significant, which gives us greater confidence in a causal interpretation for our HLM estimation.

Finally, the results of the estimation of the negative binomial model are found in table 3. Again, the coefficient estimate is in the predicted direction and statistically significant. In this case, the interpretation of the coefficient is that a 1% change in forest cover predicts a change of 0.000498 cases of ARI per household during two weeks prior to the survey. Again, while this is relatively small effect at the household level, we see that this effect is significant when aggregated across the households that the sample represents.

Taken as a whole, these results imply that standing forest reduces the incidence of respiratory health among children in India.

Valuation

In Peasah et al. (2015), the authors estimate the direct and indirect costs incurred by patients with ARI in Northern India. They describe three categories of people with ARIs: non-medically attended patients, those who utilized out-patient care, and those who utilized in-patient care. They estimate the direct and indirect costs incurred by each type of patient by age group (found in table 4). We use information on type of care in each district from the DHS data and the estimates from Peasah et al. (2015) to construct an average cost per case per district. To estimate the population of rural children under the age of five in each district, we de-normalized the DHS survey weights using the procedure recommended by DHS staff Elkasabi et al. (2020). We also adjust the costs to be in terms of 2019 dollars using the CPI published by (NOA, 2019). To account for the fact that they survey only asked respondents to recall the previous two weeks leading to the survey, we adjusted the values to yearly values. We estimated the total district level value of forests in reducing ARIs among rural children by the following formula:

$$Value_d = \mathbb{E}[\Delta Share\ ARI] \times \#\ children_d \times Mean\ case\ cost_d \times 52/2 \tag{4}$$

Across all districts in our sample, we find a mean value per square kilometer in reducing ARI incidence in rural children under the age of five of \$8922.091. This equated to a total value \$27,555,051 across all districts in the sample. This translates to \$0.76 per rural child in the districts included in our sample. This value would clearly be higher if it were to be estimated for all of India's districts with rural population.

6. Discussion

We began with the hypothesis that forests play a role in reducing the incidence of acute respiratory infections. Our estimation of a hierarchical linear model, which takes into account the nested structure of group level characteristics in our data supports our hypothesis. In particular, we estimate that each additional hectare in a district is associated with a -0.00322 change in probability that a child in our sample exhibits symptoms of ARI in the weeks leading up to the survey.

These results are limited by certain aspects of our data. Since the DHS only included details on ARI incidence in children under five, we were unable to estimate the effect of forests in mitigating ARIs in older children and adults. Also, due to the cross-sectional nature of the DHS, we could not use panel data methods to ensure that we have completely eliminated omitted variable bias. Future work could include verification of the results of this study using panel data methods on a smaller geographical sample. While the benefit transfer analysis that we conducted is a good first approximation of the value of forests of reducing ARI incidence in rural children, it greatly understates the true costs of ARI, first by only considering a subsample of the population, and secondly not accounting for the long-term effects that disease and air pollution can have on the development and lifelong earnings of children. For instance, Merfeld (2023) shows substantial losses to agricultural productivity as a result of air pollution. Azziz-Baumgartner et al. (2022) find significantly lower cognitive scores for children who had previously had an ARI. Grosse and Zhou (2021) provide a helpful overview of the lifetime monetary value of cognitive ability, and, conversely, the cost of disease which lowers cognitive ability. Additionally, deforestation is often associated with significant burning which produces airborne particulate matter, which has a negative effect on human health through channels other than ARIs. In short, our analysis should be considered as the lower bound of the value that India's forests have for reducing the incidence of ARIs.

Our work is useful in the valuation of India's forests. Correctly estimating the value of these forests is an essential tool for policymakers who should be attempting to manage these forests in a socially optimal way. When trade-offs need to be considered between economic development or resource extraction and conservation, understanding the full array of benefits that forests provide is necessary to make informed decisions.

7. Tables and Figures

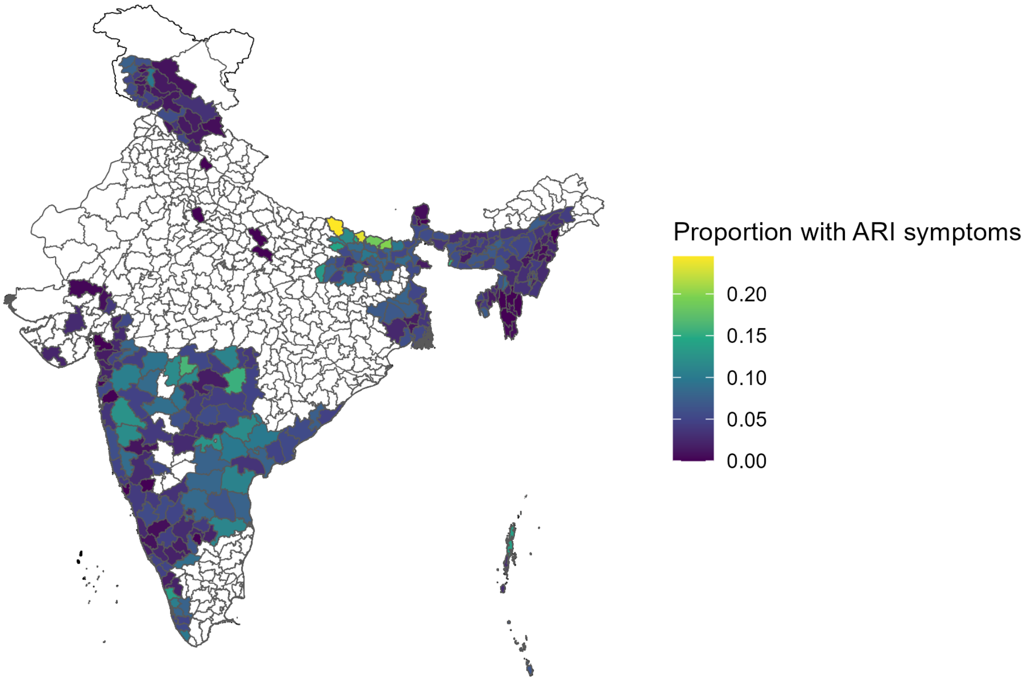

Figure 1: Proportion of sample with ARI symptoms by district

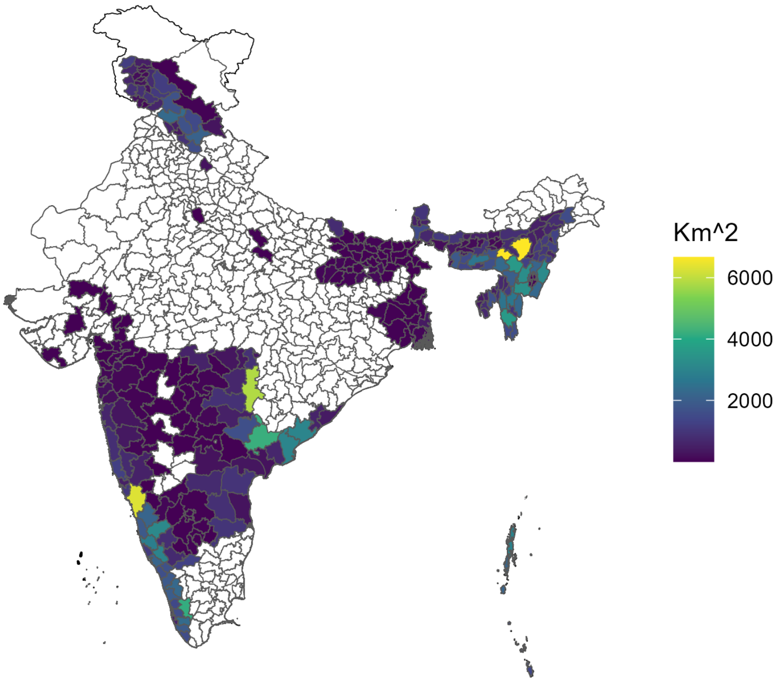

Figure 2: Square kilometers of forest

Table 1: The effect of forests in reducing the proportion of ARI symptomatic children in surveyed households.

	(1) OLS	(2) HLM
main		
Ln(District forested area km2)	-0.00273***	-0.00322***
	(0.00122)	(0.00738)
Primary	0.00220	0.00779**
	(0.72259)	(0.02373)
Secondary	-0.00839*	0.00735**
	(0.05387)	(0.01183)
Higher	-0.0224***	-0.00330
	(0.00022)	(0.45128)
Clean cooking	-0.00512	-0.00582**
	(0.16033)	(0.01259)
Smoking present in HH	0.00192	0.00186
	(0.67309)	(0.46899)
Mean child age in HH (months)	-0.000289***	-0.000113
	(0.00786)	(0.12293)
Schedule caste	-0.00202	-0.00163
	(0.72174)	(0.67728)
Schedule tribe	-0.0104*	-0.0127***
	(0.08691)	(0.00501)
OBC	0.0000183	-0.00247
	(0.99706)	(0.59309)
Don't know	-0.0150	-0.00389
	(0.35462)	(0.73084)
Altitude at cluster	-0.0000111***	-0.00000556*
	(0.00617)	(0.07908)
_		
ln_districtForestHaLag1		
Month effects	Yes	Yes
Observations	47712	47712

p-values in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 2: Control function estimation (Wooldridge, 2015). Dependent variable: proportion of children in household with ARI symptoms

	(1)	(2)
Ln(District forested area km2)	-0.00292**	-0.00203
	(-2.365)	(-1.157)
First-stage residual	-0.00073	
	(-0.476)	
Mother's education: primary	0.00169	0.01433**
	(0.285)	(2.409)
Mother's education: secondary	-0.00494	0.01271***
	(-1.186)	(2.899)
Mother's education: higher	-0.01910***	-0.00362
	(-3.276)	(-0.610)
Clean cooking	-0.00409	-0.00695*
	(-1.196)	(-1.892)
Smoking present in HH	0.00212	0.00551
	(0.474)	(1.221)
Mean child age in HH (months)	-0.00021*	-0.00015
	(-1.839)	(-1.341)
Schedule caste	-0.00062	-0.00209
	(-0.122)	(-0.396)
Schedule tribe	-0.00645	0.00537
	(-1.148)	(0.852)
OBC	0.00210	-0.00310
	(0.464)	(-0.640)
Don't know	-0.01288	-0.00522
	(-0.815)	(-0.330)
Altitude at cluster	-0.00000	0.00002***
	(-1.139)	(4.072)
First-stage residual		0.00040
		(0.177)
Month effects	Yes	Yes
State effects	No	Yes
Observations	47086	47086
Adjusted R-squared	0.00293	0.0128
AIC	-2184.4	-2634.9

Wild cluster t-statistics in parentheses.

* p<0.10, ** p<0.05, *** p<0.01

Table 3: Negative-binomial estimation. Dependent variable: count of ARI-positive children in household.

	(1) Number of cases in HH
Number of cases in HH	
Ln(District forested area km2)	-0.0498***
	(0.00003)
Primary	0.0401
	(0.59836)
Secondary	-0.0450
	(0.43640)
Higher	-0.293***
	(0.00356)
Schedule caste	-0.0326
	(0.68284)
Schedule tribe	-0.166*
	(0.07197)
OBC	0.0294
	(0.68033)
Don't know	-0.293
	(0.26163)
Altitude at cluster	-0.0000866
	(0.19528)
Clean cooking	-0.0830*
	(0.09916)
Smoking present in HH	0.00670
	(0.91330)
Mean child age in HH (months)	-0.00293*
	(0.05583)
Number of children in HH	0.459***
	(0.00000)
Constant	-2.934***
	(0.00111)
/	
lnalpha	-0.103
	(0.43054)
Month effects	Yes
Observations	47086

p-values in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4: Cost of case by type of care for children < 5 years from Peasah et al. (2015)

Category	**Direct**		**Indirect**	
	Private	Public	Private	Public
No care	$1.10	$1.10	-	-
Outpatient	$8.00	$4.90	$2.70	$9.50
Inpatient	$135	$55	$21	$24

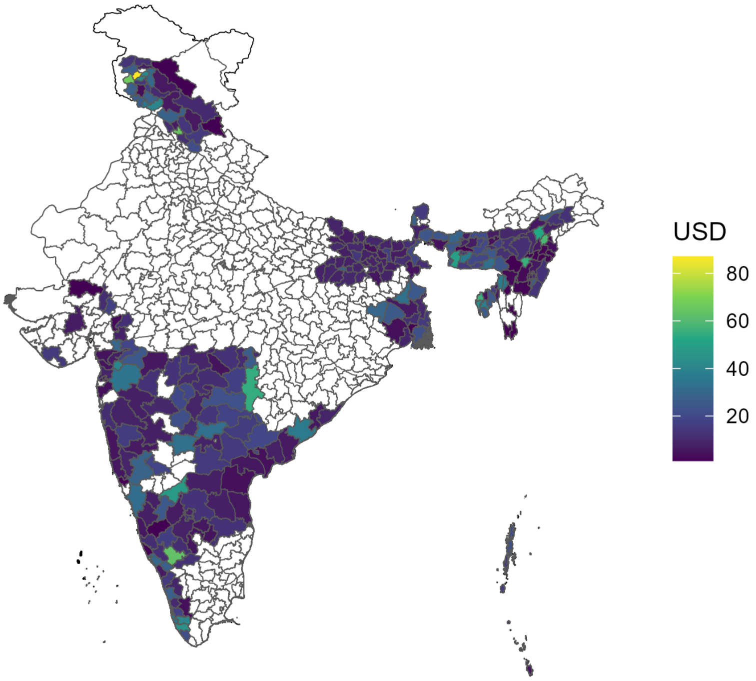

Figure 3: Average cost per ARI case by district

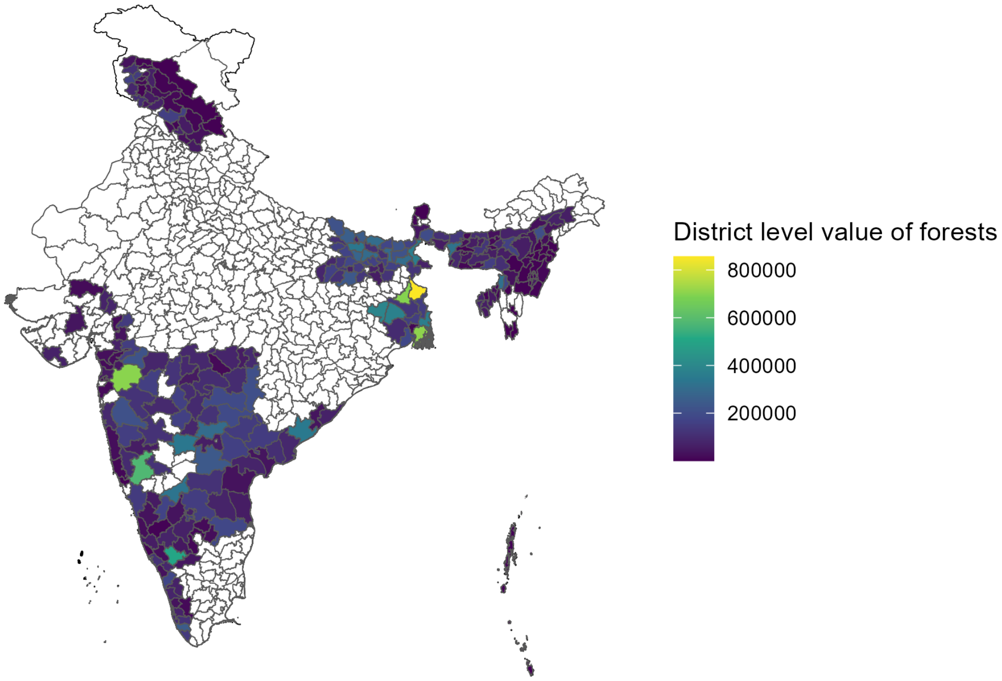

Figure 4: District yearly value of forests in reducing ARI incidence among rural children (adjusted to 2019 USD)

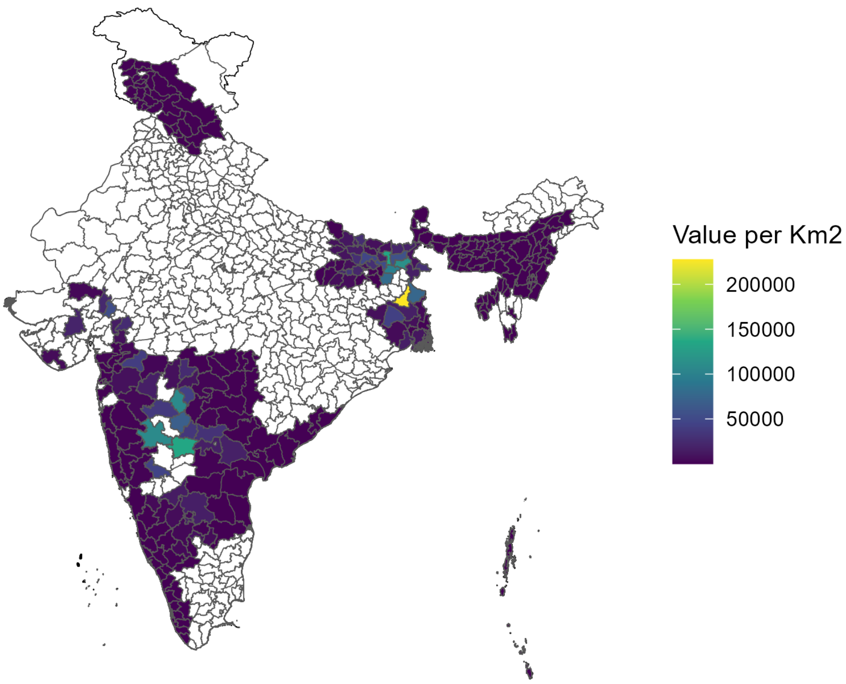

Figure 5: District yearly value of forests in reducing ARI incidence among rural children (adjusted to 2019 USD)

References

Consumer Price Index Numbers on Base 2012=100 for Rural, Urban and Combined for the Month of June 2019. Technical report, The National Statistical Office (NSO), Ministry of Statistics and Programme Implementation, 2019a. URL https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=1578589.

The DHS Program - India: Standard DHS, 2019-20, 2019b. URL https://dhsprogram.com/methodology/survey/survey-display-541.cfm.

Stata Statistical Software, 2019c.

Alberto Abadie, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. When Should You Adjust Standard Errors for Clustering?*. *The Quarterly Journal of Economics*, 138(1): 1–35, February 2023. ISSN 0033-5533. doi: 10.1093/qje/qjac038. URL https://doi.org/10.1093/qje/qjac038.

Eduardo Azziz-Baumgartner, Rosalba Gonzalez, William Davis, Arlene Calvo, Natalie Olson, Lauren Grant, Morgan Hess-Holtz, Vic Veguilla, Rafael Rauda, Susan Cornelia Kaydos-Daniels, Nestor Sosa, Evelina I. Aedo Ruíz, Julio Armero Guardado, Rachael Porter, Danilo Franco, Juan Miguel Pascale, and Georgina Peacock. Lower cognitive scores among toddlers in birth cohorts with acute respiratory illnesses, fevers, and laboratory-confirmed influenza. *Influenza and Other Respiratory Viruses*, 16(1):101–112, January 2022. ISSN 1750-2640. doi: 10.1111/irv.12904. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8692816/.

Simone C. Bauch, Anna M. Birkenbach, Subhrendu K. Pattanayak, and Erin O. Sills. Public health impacts of ecosystem change in the Brazilian Amazon. *Proceedings of the National Academy of Sciences*, 112(24):7414–7419, June 2015. doi: 10.1073/pnas.1406495111. URL https://www.pnas.org/doi/abs/10.1073/pnas.1406495111. Publisher: Proceedings of the National Academy of Sciences.

Julia Berazneva and Tanya S. Byker. Does Forest Loss Increase Human Disease? Evidence from Nigeria. *The American Economic Review*, 107(5):516–521, May 2017. ISSN 0002-8282. doi: 10.1257/aer.p20171132.

Kenneth A. Bollen, Paul P. Biemer, Alan F. Karr, Stephen Tueller, and Marcus E. Berzofsky. Are Survey Weights Needed? A Review of Diagnostic Tests in Regression Analysis. *Annual Review of Statistics and Its Application*, 3(1):375–392, 2016. doi: 10.1146/annurev-statistics-011516-012958. URL https://doi.org/10.1146/annurev-statistics-011516-012958. _eprint: https://doi.org/10.1146/annurev-statistics-011516-012958.

Javier Cortes-Ramirez, Juan D. Wilches-Vega, Olga M. Paris-Pineda, J.E. Rod, Lkhagvadulam Ayurzana, and Peter D. Sly. Environmental risk factors associated with respiratory diseases in children with socioeconomic disadvantage. *Heliyon*, 7(4):e06820, April 2021. ISSN 2405-8440. doi: 10.1016/j.heliyon.2021.e06820. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8093469/.

Mahmoud Elkasabi, Ren Ruilin, and Thomas W. Pullman. Multilevel Modeling Using DHS Surveys: A Framework To Approximate Level-Weights. Technical Report 27, ICF, Rockville, Maryland, USA, 2020. URL https://www.dhsprogram.com/pubs/pdf/MR27/MR27.pdf.

Francisco J. Escobedo, John E. Wagner, David J. Nowak, Carmen Luz De la Maza, Manuel Rodriguez, and Daniel E. Crane. Analyzing the cost effectiveness of Santiago, Chile's policy of using urban forests to improve air quality. *Journal of Environmental Management*, 86 (1):148–157, January 2008. ISSN 0301-4797. doi: 10.1016/j.jenvman.2006.11.029. URL https://www.sciencedirect.com/science/article/pii/S0301479706003823.

M. Ezzati and D. M. Kammen. Quantifying the effects of exposure to indoor air pollution from biomass combustion on acute respiratory infections in developing countries. *Environmental Health Perspectives*, 109(5):481–488, May 2001. doi: 10.1289/ehp.01109481. URL https://ehp.niehs.nih.gov/doi/10.1289/ehp.01109481. Publisher: Environmental Health Perspectives.

Scott D. Grosse and Ying Zhou. Monetary Valuation of Children's Cognitive Outcomes in Economic Evaluations from a Societal Perspective: A Review. *Children*, 8(5):352, April 2021. ISSN 2227-9067. doi: 10.3390/children8050352. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8146900/.

M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160):850–853, November 2013. doi: 10.1126/science.1244693. URL https://www.science.org/doi/10.1126/science.1244693. Publisher: American Association for the Advancement of Science.

Md Masud Hasan, Kamal Kumar Saha, Rossita Mohamad Yunus, and Khorshed Alam. Prevalence of acute respiratory infections among children in India: Regional inequalities and risk factors. *Maternal and Child Health Journal*, 26(7):1594–1602, July 2022. ISSN 1092-7875. doi: 10.1007/s10995-022-03424-3. URL https://link.springer.com/article/10.1007/s10995-022-03424-3. Place: New York Publisher: Springer/Plenum Publishers WOS:000784118500002.

J. A. Hausman. Specification Tests in Econometrics. *Econometrica*, 46(6):1251–1271, 1978. ISSN 0012-9682. doi: 10.2307/1913827. URL https://www.jstor.org/stable/1913827. Publisher: [Wiley, Econometric Society].

Antje Inness, Melanie Ades, Anna Agustí-Panareda, Jérôme Barré, Anna Benedictow, Anne-Marlene Blechschmidt, Juan Jose Dominguez, Richard Engelen, Henk Eskes, Johannes Flemming, Vincent Huijnen, Luke Jones, Zak Kipling, Sebastien Massart, Mark Parrington, Vincent-Henri Peuch, Miha Razinger, Samuel Remy, Michael Schulz, and Martin Suttie. The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, 19(6):3515–3556, March 2019. ISSN 1680-7316. doi: 10.5194/acp-19-3515-2019. URL https://acp.copernicus.org/articles/19/3515/2019/. Publisher: Copernicus GmbH.

Pamela Jagger and Gerald Shively. Land use change, fuel use and respiratory. health in Uganda. *Energy Policy*, 67:713–726, April 2014. ISSN 0301-4215. doi: 10.1016/j.enpol.2013.11.068. URL https://linkinghub.elsevier.com/retrieve/pii/S0301421513011981. Place: Oxford Publisher: Elsevier Sci Ltd WOS:000332815300070.

Taejoon Kim, Bokyeong Song, Kyoung Sang Cho, and Im-Soon Lee. Therapeutic Potential of Volatile Terpenes and Terpenoids from Forests for Inflammatory Diseases. *International Journal of Molecular Sciences*, 21(6):2187, January 2020. ISSN 1422-0067. doi: 10.3390/

ijms21062187. URL https://www.mdpi.com/1422-0067/21/6/2187. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

Peter S. Larson, Leon Espira, Bailey E. Glenn, Miles C. Larson, Christopher S. Crowe, Seoyeon Jang, and Marie S. O'Neill. Long-Term PM2.5 Exposure Is Associated with Symptoms of Acute Respiratory Infections among Children under Five Years of Age in Kenya, 2014. *International Journal of Environmental Research and Public Health*, 19(5):2525, February 2022. ISSN 1661-7827. doi: 10.3390/ijerph19052525. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8909525/.

J. Lee, B. J. Park, Y. Tsunetsugu, T. Ohira, T. Kagawa, and Y. Miyazaki. Effect of forest bathing on physiological and psychological responses in young Japanese male subjects. *Public Health*, 125(2):93–100, February 2011. ISSN 0033-3506. doi: 10.1016/j.puhe.2010.09.005. URL https://www.sciencedirect.com/science/article/pii/S0033350610003203.

Jessica J. Lewis, John W. Hollingsworth, Ryan T. Chartier, Ellen M. Cooper, William Michael Foster, Genna L. Gomes, Peter S. Kussin, John J. MacInnis, Bijaya K. Padhi, Pinaki Panigrahi, Charles E. Rodes, Ian T. Ryde, Ashok K. Singha, Heather M. Stapleton, Jonathan Thornburg, Cora J. Young, Joel N. Meyer, and Subhrendu K. Pattanayak. Biogas Stoves Reduce Firewood Use, Household Air Pollution, and Hospital Visits in Odisha, India. *Environmental Science & Technology*, 51(1):560–569, January 2017. ISSN 0013-936X. doi: 10.1021/acs.est.6b02466. URL https://doi.org/10.1021/acs.est.6b02466. Publisher: American Chemical Society.

J. Scott Long. *Regression Models for Categorical and Limited Dependent Variables*, volume 7 of *Advanced Quantitative Techniques in the Social Sciences*. Sage Publications, Inc., March 2024. URL https://us.sagepub.com/en-us/nam/regression-models-for-categorical-and-limited-dependent-variables/book6071.

Alan D Lopez, Colin D Mathers, Majid Ezzati, Dean T Jamison, and Christopher JL Murray. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *The Lancet*, 367(9524):1747–1757, May 2006. ISSN 0140-6736. doi: 10.1016/S0140-6736(06)68770-9. URL https://www.sciencedirect.com/science/article/pii/S0140673606687709.

Xiaoxiu Lun, Ying Lin, Fahe Chai, Chong Fan, Hong Li, and Junfeng Liu. Reviews of emission of biogenic volatile organic compounds (BVOCs) in Asia. *Journal of Environmental Sciences*, 95:266–277, September 2020. ISSN 1001-0742. doi: 10.1016/j.jes.2020.04.043. URL `https://www.sciencedirect.com/science/article/pii/S1001074220301923`.

Amanuel Mengistu Merera. Determinants of acute respiratory infection among under-five children in rural Ethiopia. *BMC Infectious Diseases*, 21(1):1203, November 2021. ISSN 1471-2334. doi: 10.1186/s12879-021-06864-4. URL `https://doi.org/10.1186/s12879-021-06864-4`.

Joshua D. Merfeld. Air Pollution and Agricultural Productivity in a Developing Country. *SSRN Electronic Journal*, 2023. ISSN 1556-5068. doi: 10.2139/ssrn.4483694. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4483694`.

Daniel C. Miller, Pushpendra Rana, and Catherine Benson Wahlén. A Crystal Ball for Forests?: Analyzing the Social-Ecological Impacts of Forest Conservation and Management over the Long Term. *Environment and Society*, 8(1):40–62, January 2017. ISSN 2150-6779, 2150-6787. doi: 10.3167/ares.2017.080103. URL `https://www.berghahnjournals.com/view/journals/environment-and-society/8/1/ares080103.xml`. Publisher: Berghahn Journals Section: Environment and Society.

David J. Nowak, Daniel E. Crane, Jack C. Stevens, and Myriam Ibarra. Brooklyn's urban forest. Technical Report NE-GTR-290, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA, 2002. URL `https://www.fs.usda.gov/treesearch/pubs/3174`.

David J. Nowak, Satoshi Hirabayashi, Allison Bodine, and Eric Greenfield. Tree and forest effects on air quality and human health in the United States. *Environmental Pollution*, 193: 119–129, October 2014. ISSN 0269-7491. doi: 10.1016/j.envpol.2014.05.028. URL `https://www.sciencedirect.com/science/article/pii/S0269749114002395`.

José Luis Montiel Olea and Carolin Pflueger. A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*, 31(3):358–369, July 2013. ISSN 0735-0015, 1537-2707. doi:

10.1080/00401706.2013.806694. URL http://www.tandfonline.com/doi/abs/10.1080/00401706.2013.806694.

Subhrendu Pattanayak, Shubhayu Saha, Pravash Sahu, Erin Sills, Ashok Singha, and JuiChen Yang. Mine over matter? Health, wealth and forests in a mining area of Orissa. *Indian Growth and Development Review*, 3(2):166–185, January 2010. ISSN 1753-8254. doi: 10.1108/17538251011084473. URL https://doi.org/10.1108/17538251011084473. Publisher: Emerald Group Publishing Limited.

Subhrendu K. Pattanayak and Alexander Pfaff. Behavior, Environment, and Health in Developing Countries: Evaluation and Valuation. *Annual Review of Resource Economics*, 1(1):183–217, October 2009. ISSN 1941-1340, 1941-1359. doi: 10.1146/annurev.resource.050708.144053. URL https://www.annualreviews.org/doi/10.1146/annurev.resource.050708.144053.

Samuel K. Peasah, Debjani Ram Purakayastha, Parvaiz A. Koul, Fatima S. Dawood, Siddhartha Saha, Ritvik Amarchand, Shobha Broor, Vaibhab Rastogi, Romana Assad, Kaisar Ahmed Kaul, Marc-Alain Widdowson, Renu B. Lal, and Anand Krishnan. The cost of acute respiratory infections in Northern India: a multi-site study. *BMC Public Health*, 15(1):330, April 2015. ISSN 1471-2458. doi: 10.1186/s12889-015-1685-6. URL https://doi.org/10.1186/s12889-015-1685-6.

Danny Pfeffermann and Michail Sverchkov. Small-Area Estimation under Informative Probability Sampling of Areas and within the Selected Areas. *Journal of the American Statistical Association*, 102(480):1427–1439, 2007. ISSN 0162-1459. URL https://www.jstor.org/stable/27639992. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Thomas Pienkowski, Borame Lee Dickens, Haoyang Sun, and Luis Roman Carrasco. Empirical evidence of the public health benefits of tropical forest conservation in Cambodia: a generalised linear mixed-effects model analysis. *The Lancet Planetary Health*, 1(5):e180–e187, August 2017. ISSN 2542-5196. doi: 10.1016/S2542-5196(17)30081-5. URL https://www.sciencedirect.com/science/article/pii/S2542519617300815.

Aleksandra Ratajczak, Artur Badyda, Piotr Oskar Czechowski, Adam Czarnecki, Michał Dubrawski, and Wojciech Feleszko. Air Pollution Increases the Incidence of Upper Respiratory Tract Symptoms among Polish Children. *Journal of Clinical Medicine*, 10(10): 2150, May 2021. ISSN 2077-0383. doi: 10.3390/jcm10102150. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8156299/.

Stephen W. Raudenbush and Anthony S. Bryk. *Hierarchical Linear Models*, volume 1. University of Chicago, 2002. URL https://us.sagepub.com/en-us/nam/hierarchical-linear-models/book9230.

David Roodman, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb. Fast and wild: Bootstrap inference in Stata using boottest. *The Stata Journal*, 19(1):4–60, March 2019. ISSN 1536-867X. doi: 10.1177/1536867X19830877. URL https://doi.org/10.1177/1536867X19830877. Publisher: SAGE Publications.

Eric W. Sanderson, Dale G. Miquelle, Kim Fisher, Abishek Harihar, Chris Clark, Jesse Moy, Peter Potapov, Nathaniel Robinson, Lucinda Royte, Dustin Sampson, Jamie Sanderlin, Charles B. Yackulic, Michael Belecky, Urs Breitenmoser, Christine Breitenmoser-Würsten, Pranav Chanchani, Stuart Chapman, Arpit Deomurari, Somphot Duangchantrasiri, Elisa Facchini, Thomas N. E. Gray, John Goodrich, Luke Hunter, Matthew Linkie, Willy Marthy, Akchousanh Rasphone, Sugoto Roy, Detrit Sittibal, Tshering Tempa, Mayuree Umponjan, and Karen Wood. Range-wide trends in tiger conservation landscapes, 2001 - 2020. *Frontiers in Conservation Science*, 4, December 2023. ISSN 2673-611X. doi: 10.3389/fcosc.2023.1191280. URL https://www.frontiersin.org/articles/10.3389/fcosc.2023.1191280. Publisher: Frontiers.

Joseph O. Sexton, Praveen Noojipady, Xiao-Peng Song, Min Feng, Dan-Xia Song, Do-Hyung Kim, Anupam Anand, Chengquan Huang, Saurabh Channan, Stuart L. Pimm, and John R. Townshend. Conservation policy and the measurement of forests. *Nature Climate Change*, 6(2):192–196, February 2016. ISSN 1758-6798. doi: 10.1038/nclimate2816. URL https://www.nature.com/articles/nclimate2816. Number: 2 Publisher: Nature Publishing Group.

Gary Solon, Steven J. Haider, and Jeffrey M. Wooldridge. What Are We Weighting For?

Journal of Human Resources, 50(2):301–316, March 2015. ISSN 0022-166X, 1548-8004. doi: 10.3368/jhr.50.2.301. URL https://jhr.uwpress.org/content/50/2/301. Publisher: University of Wisconsin Press Section: Symposium on Empirical Methods.

E. Somanathan, R. Prabhakar, and Bhupendra Singh Mehta. Decentralization for cost-effective conservation. *Proceedings of the National Academy of Sciences*, 106(11):4143–4147, March 2009. doi: 10.1073/pnas.0810049106. URL https://www.pnas.org/doi/10.1073/pnas.0810049106. Publisher: Proceedings of the National Academy of Sciences.

Jesty Saira Varghese and T. Muhammad. Prevalence, potential determinants, and treatment-seeking behavior of acute respiratory infection among children under age five in India: Findings from the National Family Health Survey, 2019-21. *BMC Pulmonary Medicine*, 23(1):195, June 2023. ISSN 1471-2466. doi: 10.1186/s12890-023-02487-4. URL https://doi.org/10.1186/s12890-023-02487-4.

Global Forest Watch. Tree Cover Loss in India. URL https://www.globalforestwatch.org/dashboards/country/IND.

Ye Wen, Qi Yan, Yangliu Pan, Xinren Gu, and Yuanqiu Liu. Medical empirical research on forest bathing (Shinrin-yoku): a systematic review. *Environmental Health and Preventive Medicine*, 24(1):70, December 2019. ISSN 1347-4715. doi: 10.1186/s12199-019-0822-8. URL https://doi.org/10.1186/s12199-019-0822-8.

Jeffrey M. Wooldridge. Control Function Methods in Applied Econometrics. *The Journal of Human Resources*, 50(2):420–445, 2015. ISSN 0022-166X. URL https://www.jstor.org/stable/24735991. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System].

Han Yu, Shanhe Jiang, and Kenneth C. Land. Multicollinearity in hierarchical linear models. *Social Science Research*, 53:118–136, September 2015. ISSN 0049-089X. doi: 10.1016/j.ssresearch.2015.04.008. URL https://www.sciencedirect.com/science/article/pii/S0049089X15000885.

8. Appendix: Additional Tables and Figures

Table 5: Variables and expected relationship to ARI

Variable	**(+/-)**	**Mechanism**
Forests	(-)	Forests filter particulate matter (PM2.5) from the air. PM2.5 exacerbates the severity of respiratory diseases. Time spent in forests is also correlated with immune system health.
Forest loss	(+)	In addition to the loss of air filtering, forest loss can be a result of burning or use of fuelwood, both which can reduce air quality.
Education of mother	(-)	Education correlated with increased access to information on treatment and prevention of ARI.
Clean cooking fuel	(-)	Use of bio-fuels is a major cause of ARI in developing countries.
Smoking in HH	(+)	Smoking degrades HH air quality, increasing incidence of and exposure to ARI.
Child's age (months)	(-)	Younger children are more susceptible to ARIs.

Table 6: Results of test for the necessity of the inclusion of survey weights suggested by Pfeffermann and Sverchkov (2007). The statistical significance of the ARI variable suggests the need for inclusion of survey weights.

	(1) Survey weights	
ARI symptoms	0.0972***	(0.00000)
Hectares of forest in district: 2018	-0.00000111***	(0.00000)
No education	0	(.)
Primary	-0.0252**	(0.02995)
Secondary	-0.00336	(0.69929)
Higher	0.0681***	(0.00000)
No clean cooking	0	(.)
Clean cooking	0.0505***	(0.00000)
Schedule caste	-0.0211*	(0.06999)
Schedule tribe	-0.495***	(0.00000)
OBC	-0.0910***	(0.00000)
None of them	0	(.)
Male	0	(.)
Female	-0.00236	(0.72573)
child's age in months	0.0000482	(0.80522)
Altitude at cluster	-0.000385***	(0.00000)
Month of survey=6	0	(.)
Month of survey=7	0.361***	(0.00000)
Month of survey=8	0.110	(0.14000)
Month of survey=9	0.0992	(0.18426)
Month of survey=10	0.108	(0.14830)
Month of survey=11	-0.0451	(0.54755)
Month of survey=12	-0.0863	(0.25695)
No smoking in HH	0	(.)
Smoking present in HH	-0.208***	(0.00194)
Constant	1.214***	(0.00000)
Observations	67963	

p-values in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 7: Results of the Hausman (1978) test for the necessity of the inclusion of survey weights. This provides some evidence that survey weights are necessary to control for endogenous sampling.

		(1) Survey weights
ARI symptoms		
Hectares of forest in district: 2018	-4.60e-08***	(0.00061)
No education	0	(.)
Primary	0.00235	(0.65271)
Secondary	-0.00449	(0.23903)
Higher	-0.0175***	(0.00176)
No clean cooking	0	(.)
Clean cooking	-0.00458	(0.14371)
Schedule caste	-0.00294	(0.56555)
Schedule tribe	-0.0118**	(0.02987)
OBC	-0.000586	(0.89773)
None of them	0	(.)
Male	0	(.)
Female	-0.00940***	(0.00055)
child's age in months	-0.000574***	(0.00000)
Altitude at cluster	-0.00000549	(0.13706)
Constant	0.0983***	(0.00000)
Observations	67963	

p-values in parentheses

Note: smoking in household and month of survey variables excluded due to lack of full rank of the differenced variance matrix with their inclusion.

* $p < .10$, ** $p < .05$, *** $p < .01$

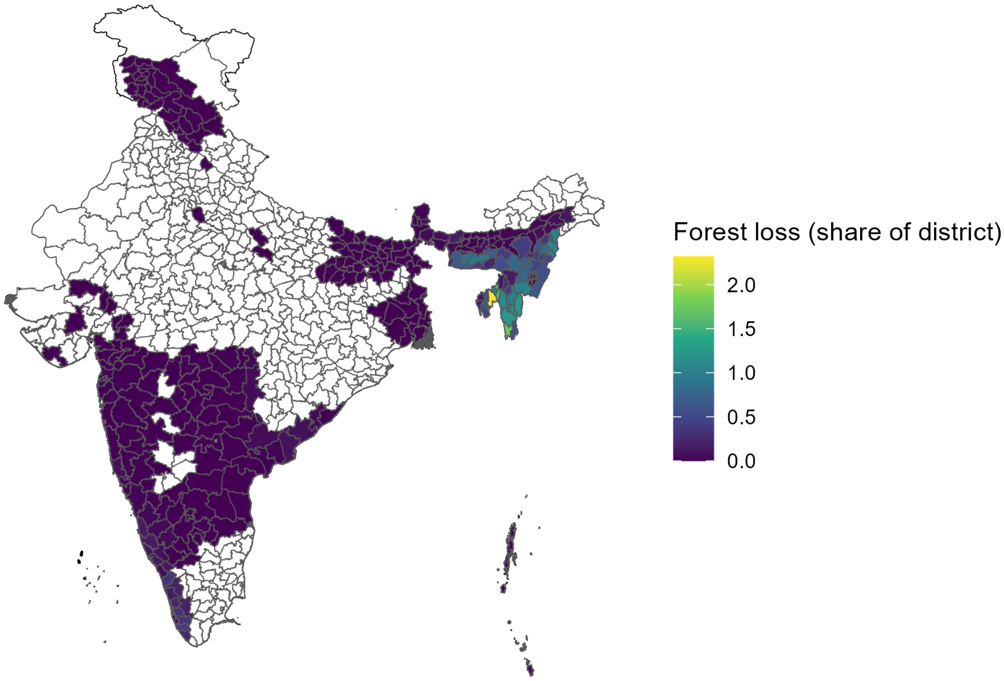

Figure 6: Forest loss in year prior to survey (% of district area)

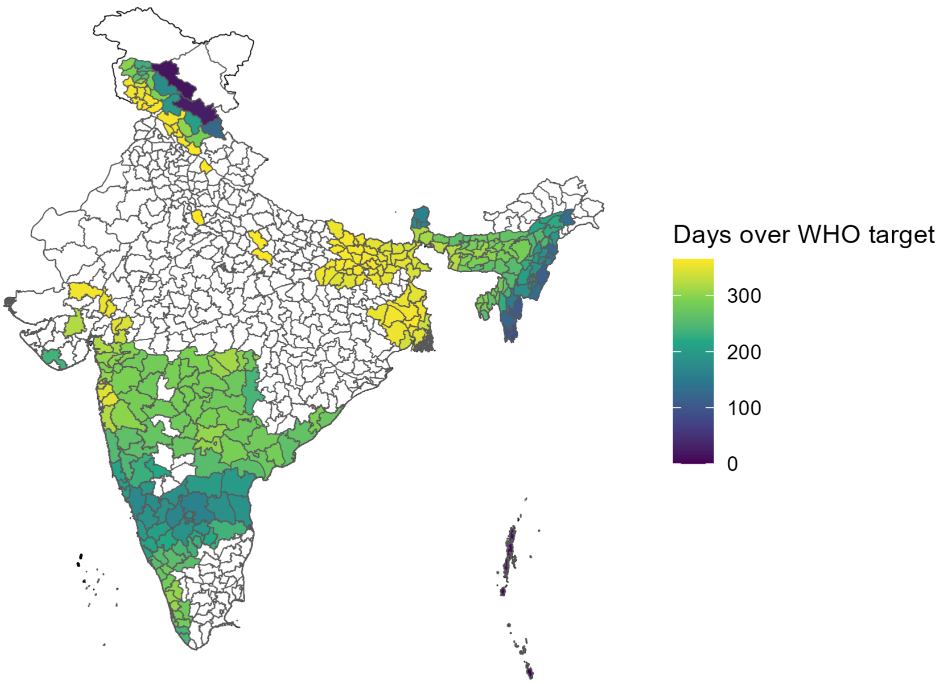

Figure 7: PM2.5 days over WHO interim target