WP 2020-04
May 2020

Working Paper

Department of Applied Economics and Management
Cornell University, Ithaca, New York 14853-7801 USA

COOKING FUEL CHOICE, INDOOR AIR QUALITY AND CHILD MORTALITY IN INDIA

Arnab K. Basu, Tsenguunjav Byambasuren, Nancy H. Chau, and Neha Khanna

Cooking Fuel Choice, Indoor Air Quality and Child Mortality in India*

Arnab K. Basu, Tsenguunjav Byambasuren, Nancy H. Chau and Neha Khanna†

This version: May 2020

Abstract: Indoor air pollution (IAP)–predominantly from the use of solid fuel for cooking–is a global health threat, particularly for women and young children, and one of the leading causes of infant deaths worldwide in developing countries. We estimate the causal effect of cooking fuel choice on infant mortality in India, focusing on children under five years of age using pooled cross-sectional data from the National Family Health Survey (NFHS) over the period 1992–2016. To address the potential endogeneity in the relationship between fuel choice and mortality, we instrument for cooking fuel choice using a speed of change in forest cover and ownership status of agricultural land, which induce significant variations in fuel type. We find that cooking fuel choice has a statistically significant impact on under-five and neonatal mortality, raising the mortality risk by 4.9 percent. We also find that the past literature has overestimated the association between under-five mortality and polluting fuel use by about 0.6 percentage points or equivalently, 152,000 deaths per year nationally. Our result is robust to a set of alternative specifications with the inclusion of various controls and different estimation strategies.

JEL Classification: I18, N35, Q53.

Keywords: cooking fuel, indoor air pollution, infant mortality, India.

*We have benefited from comments by Ali Abbas, Chris Barrett, Andrew Foster, Maulik Jagnani, Gary Lin, Hemant Pullabhotla, Avralt-Od Purevjav and participants at the Northeast Universities Development Economics Consortium (NEUDC) Conference at Northwestern University (2019). The usual disclaimer applies.

†Basu: Charles H. Dyson School of Applied Economics and Management, Cornell University, 441 Warren Hall, Ithaca, NY 14853 & Institute for the Study of Labor (IZA) (email: arnab.basu@cornell.edu); Byambasuren: Charles H. Dyson School of Applied Economics and Management, Cornell University, 424 Warren Hall, Ithaca, NY 14853 (email: tb497@cornell.edu); Chau: Charles H. Dyson School of Applied Economics and Management, Cornell University, 201A Warren Hall, Ithaca, NY 14853 & Institute for the Study of Labor (IZA) (email: hyc3@cornell.edu); Khanna: Department of Economics and Environmental Studies Program, Binghamton University, Library Tower 1004, Binghamton, NY 13902-6000 (email: nkhanna@binghamton.edu).

1 Introduction

Indoor air pollution (IAP) is produced mainly by incomplete combustion of polluting fuels used for cooking, heating, and lighting and remains the single largest environmental health risk factor worldwide (WHO, 2016). Almost three billion people—41% of the world's population—have been using open fire or simple stoves fueled by kerosene, coal, wood, animal dung and crop waste for cooking and as domestic sources of energy for the past three decades (WHO, 2018b).

Around 95% of these people live in poverty in low and middle-income countries of Southeast Asia, Western Pacific, and Africa: 80% of the population in China, 82% in India, 87% in Ghana, 95% in Afghanistan, and 95% in Chad rely primarily on polluting cooking fuels (Duflo et al., 2008b). The combination of traditional cooking stoves and polluting fuels generates high levels of hazardous indoor air pollutants. Each year, close to 4 million people die because of diseases attributable to indoor air pollution (including heart disease, respiratory disease, stroke, and cancer) caused by an inefficient use of polluting fuels for cooking and heating (WHO, 2018b).[1]

Since women are mainly responsible for cooking, and children spend most of their time with their mothers in developing countries, women and young children (especially, children under five years of age) tend to be more exposed to IAP (Edwards and Langpap, 2012). In India, approximately 56% of under-five children stay with their mothers at all times including during cooking (Rehfuess et al., 2011; Martin et al., 2014). Thus, the environmental risks from IAP to health are highest among the most vulnerable members of society. IAP from cooking with solid fuels is the biggest cause of disability-adjusted life years (DALYs)[2] lost in Southeast Asia and Sub-Saharan Africa, and the third leading cause of DALYs lost globally (Apte and Salvi, 2016). Moreover, acute lower respiratory infections (ALRI), including pneumonia is the second dominant cause of deaths in children under five years of age in the

[1] In 2016, IAP from solid fuel use resulted in 3.8 million premature deaths, equivalent to 6.7% of global mortality, greater than the toll due to malaria, tuberculosis and HIV/AIDS combined. Of these deaths, 403,000 were among children under 5 years of age (WHO, 2016, 2018a). Air pollution is the leading environmental factor for death in India, accounting for about 1.2 million deaths in 2017, nearly 40 percent of which are due to poor indoor air quality (Global Burden of Disease 2017).

[2] The DALY is the most commonly used measure of national burden of disease and combines the years of life lost due to disability with the years of life lost due to death.

world after prematurity, and one-third of ALRI-related deaths are because of poor quality air inside the house (WHO, 2000, 2018c).[3]

There has been a growing literature focusing on the health impacts of exposure to IAP. Nearly 200 publications, most of which are epidemiological studies, reported health effects of solid fuel combustion in Chinese households and documented strong evidence for adverse health outcomes including chronic obstructive pulmonary disease (COPD), ALRI, asthma, lung cancer, and immune system impairment (Zhang and Smith, 2007). One of the earliest works which investigate the health impact of IAP found a high correlation between using a traditional stove and having symptoms of respiratory illness using a linear probability model with a variety of controls (Duflo et al., 2008a).

The first randomized control trial (RCT) experiments on health effects of IAP were conducted in the city of San Marcos, Guatemala, by Diaz et al. (2007) and Smith-Sivertsen et al. (2009). Using logistic random intercept models, they found that the use of improved cooking stoves (*planchas*) has a protective health effect by reducing exposure to IAP and symptoms of headache and sore eyes during 18 months of their follow-up. Imelda (2018) evaluated the causal effect of the Indonesian government program of subsidizing households to switch from kerosene to liquid petroleum gas (LPG) as the cooking fuel choice using a quasi-experimental approach. Employing a difference-in-differences (DID) method, she found that the program led to a 1.1 percent reduction in infant mortality rate.[4] Hanna

[3]Relatedly, Silwal and McKay (2015) find that the use of firewood instead of kerosene, LPG and electricity for cooking damages the individual's lung capacity by 9.4 percent in Indonesia using proximity to the nearest market as an instrument for household fuel choice. In addition, Edwards and Langpap (2012) investigate the impacts of firewood consumption and whether mother cooks while caring for children on children's respiratory health in Guatemala using household's gas stove ownership and mother's age as IVs for the two regressors of interest, and suggest that their key explanatory variables cause incidences of respiratory symptoms among under-five children. It is worth noting that the validation of the exclusion restriction and exogeneity of instruments utilized in these two studies are weakly justified. Pitt et al. (2006) also examine the effect of time spent cooking on incidence of any respiratory symptom for all adults and adult women based on an instrumental variable strategy, considering that there is heterogeneity bias and measurement error in the time spent cooking. Using gender-specific hierarchies as instruments for exposure to IAP, they suggest that a four hour per day increase in the time spent cooking leads to a 10.8 percentage point increase in the probability of having a respiratory symptom. It is also important to note that the Silwal and McKay (2015), Edwards and Langpap (2012) and Pitt et al. (2006) papers have a different focus in terms of the research questions, and are thus distinct from ours.

[4]A related work is Imelda (2018) who estimates the causal effect of IAP (proxied by a household fuel-switching program) on infant mortality using a difference-in-differences (DID) estimation strategy on the effect of a kerosene (polluting fuel) to liquid petroleum gas (LPG–clean fuel) conversion program implemented by the Indonesian government. In a follow-up paper, Imelda (2019) also estimates 3.3 percentage points

et al. (2016) conduct the largest RCT with a 4-year of follow-up in rural Orissa, India to address the long term impacts of improved cookstoves. The study provides evidence that improved cookstoves did not reduce smoke exposure following the second year of installation, or improve health of recipients and greenhouse gas emissions at all because they were not used regularly and recipients did not maintain them properly.

While the extant literature has mainly focused on the effectiveness of specific policies and programs (e.g., improved cooking stoves, house construction, and voucher allocation for electrification) on reducing IAP and improving selected health outcomes (Bruce et al., 2004; Duflo et al., 2008a; Smith-Sivertsen et al., 2009; Hanna et al., 2016; Barron and Torero, 2017), Naz et al. (2016) has explored the the effect of IAP on under-five mortality in India. However, Naz et.al.'s analysis of the odds ratio of 1.30 between polluting fuel use and under-five mortality ignores potential omitted variables bias and endogeneity issues. For example, Naz et.al. does not account for factors such as number of people in the household, small size of dwelling, and other regional-level demographic and environmental factors (e.g., characteristics of local governors and administrators, culture of the region, distance to the nearest metropolitan areas, and ambient air and soil quality)[5] that simultaneously affect IAP concentration and mortality.

Our contribution to this nascent literature on the IAP-child mortality link thus lies in estimating the causal impact of indoor air pollution through the use of polluting fuels for cooking on under-five mortality after accounting for the reverse causality between mortality and cooking fuel choice. The issue of reverse causality, from mortality (health outcome) to

reduction in infant mortality rate in response to a 10 percentage points increase in the intensity of the same government program by replacing a binary treatment variable with a continuous variable of program intensity in her DID equation.

[5]For instance, the demographic and political characteristics of the state governors, such as their age, gender, education, political power and affiliation, and relationship with the government, could affect the implementation of the central government policy and local policy initiation about household fuel use and air pollution in general. It is also reasonable to consider that there exist region-specific socio-cultural trends that prevent households from switching to modern and clean energy. Many people in developing countries, who live in rural areas, tend to prefer to use animal dung not just due to its abundance but because they believe that it is clean and natural. The proximity to the nearest town and city represents access to the socioeconomic and medical resources, while it also proxies the closeness to the clean fuels, including electricity and various types of gases, and remoteness of the polluting fuels such as firewood and animal dung. Outdoor air quality also reflects, for example, the abundance of forests and coal deposits, both of which potentially determine household fuel decisions, while soil quality could be a proxy of arable land that possibly demonstrates the availability of agricultural crop waste.

cooking fuels, is important for several reasons. First, switching to a cleaner cooking fuel is a readily available means to prevent another mortality case immediately subsequent to a mortality case in a household. Second, air pollution can also adversely affect an individual's long-term earnings through poor health and low productivity (Graff Zivin and Neidell, 2013; Isen et al., 2017). Relatively poorer households are caught in a vicious cycle (or poverty trap) wherein they are only able to afford cheaper and more polluting cooking fuel options, which adversely affects household health and mortality and, in turn, household earnings (Hanna and Oliva, 2015; Graff Zivin and Neidell, 2012, 2018; Chang et al., 2016, 2019).[6]

We use two instrumental variables for household fuel choice–forest cover and agricultural land ownership–to address this potential endogeneity. Density of forest cover across different locations determines the availability or access, lower opportunity costs for households to collect, and lower prices for local firewood (often classified as a polluting fuel). Furthermore, households that own land for agricultural purposes are more likely to use polluting fuels such as agricultural crop waste, animal dung, and even firewood. We thus provide the first empirical estimate of the causal effect of IAP, as determined by cooking fuel choice, on under-five mortality by relying on plausibly exogenous variations in IAP introduced by the speed of change in forest cover and status of agricultural land ownership. Conditional on other controls included in our empirical specifications, we do not expect these two variables to have any impact on child mortality.

Our analysis is based on a large-scale household survey collected throughout India that recorded the health and demographic information, including type of fuels used for cooking, from 1992 to 2016. Specifically, we rely primarily on three rounds of India's National Family Health Survey (NFHS) — NFHS-1 (1992–93), NFHS-2 (1998–99), and NFHS-4 (2015–16) which include detailed observations on 369,416 singleton live-born children, of whom 19,517 died in the 5-years prior to the respective survey years. Our analysis shows that a household using solid fuels for cooking has a 4.9 percent higher probability of experiencing under-five child mortality. Furthermore, we re-estimate the existing models in the literature with

[6]As an example, a strong negative effect of air pollution (carbon monoxide–CO) on fourth-grade test scores (math and language skills) was observed in Santiago, Chile, and a 50% increase in CO in Santiago between 1990 to 2005 reduced an individual's lifetime earnings by around US$100 million (Bharadwaj et al., 2017).

additional controls for exposure to pollution within the household and find find that the use of dirty fuel for cooking is associated with an increased risk of under-five mortality by 0.8 percentage points. This is lower than the previous estimates reported in the literature by about 0.6 percentage points. These earlier estimates, on the consequences of IAP, have been questioned in terms of their reliability due to inadequate controls for health outcomes and lack of convincing identification strategies (Duflo et al., 2008b). We assess the robustness of our findings by estimating a variety of specifications with additional controls and fixed effects.

Taken together, this paper makes the following two contributions to the literature on the impact of IAP on child mortality. First, to our knowledge, this is the first attempt to empirically estimate the causal effect of IAP, as proxied by cooking fuel choice, on infant mortality while addressing the endogeneity in the relationship between cooking fuel choices and mortality. Existing studies mostly based on epidemiological estimates of the IAP-mortality relationship rather than causal estimates (Duflo et al., 2008b). While the endogeneity issue in the mortality-IAP (or -cooking fuel) relationship has been recognized (Schindler et al., 2017), it has not been addressed in empirical settings yet, perhaps due to the challenge in finding a valid instrumental variable.[7]

Second, we utilize the NFHS (also called Demographic and Health Survey–DHS) — a widely-accepted gold standard for research in the developing world — datasets that cover 601,509 representative households from all 36 states and 640 districts of India over the last 25 years. Most of the papers that study indoor air quality in developing countries largely focused on a rural village of Orissa in India (Duflo et al., 2008a; Hanna et al., 2016), the city of San Marcos in Guatemala (Smith-Sivertsen et al., 2009), and the rural village of La Victoria in the western highlands of Guatemala (Bruce et al., 2004). A detailed and large-scale dataset collected from this nationwide household survey, covering both urban and rural areas, allows us to provide more broadly representative empirical estimates of the causal relationship between cooking fuel choice, and therefore IAP, and infant mortality. We consider a total of 12 types of cooking fuels including kerosene, coal/lignite, charcoal, wood,

[7]Those few studies mentioned earlier attempt to address the endogeneity in the lung capacity-use of firewood, respiratory illness-time spent cooking, and children's respiratory health-firewood consumption relationships using IV strategy.

straw/shrubs/grass, agricultural crop waste, and animal dung as a dirty fuel, and electricity, LPG, natural gas and biogas as a clean fuel. In addition, we consider mortality of four different age-groups including neonatal, post-neonatal, child, and under-five.

The remainder of the paper is organized as follows. Section 2 presents the background on IAP and child mortality in India and provides the trend analysis of under-five mortality attributed to the cooking fuel types. Section 3 lays out the empirical strategy, and Section 4 describes the data and presents descriptive statistics for the sample. Section 5 presents model results and a set of robustness tests. Section 6 concludes.

2 Background

With a population of 1.4 billion, India is the second-most populous country in the world and the tenth-biggest contributor to global gross domestic product. Over 72% of households in India (more than 90% of the rural population and 31% of the urban population) use solid fuels as a primary source of energy and for cooking. In this section, we first discuss India's challenges related to IAP due to cooking fuel choice and the potential effect on early childhood (under-five) mortality. We then present and discuss the trends in India's under-five mortality incidence in relation to type of cooking fuels.

2.1 Indoor Air Pollution and Infant Mortality in India

The United States Environmental Protection Agency (EPA) sets standards for PM_{10} concentrations at 50 $\mu g/m^3$ based on an annual average, and at 150 $\mu g/m^3$ based on a 24-hour average (`https://www3.epa.gov/region1/airquality/pm-aq-standards.html`). However, the 24-hour average of PM_{10} concentration in solid fuel firing households in India often exceeds 2,000 $\mu g/m^3$ (Smith, 2000). Menon (1988) and Saksena et al. (1992) found higher concentrations of PM_{10} (20,000 $\mu g/m^3$) near the cooking location in India, with the concentration decreasing substantially with distance away from kitchen.

According to the World Health Organization (WHO), 3.5% of the total burden of disease in India has been caused by IAP (Bonjour et al., 2007), while 20% of deaths among children aged under-five can be attributed to the use of solid fuels (Bassani et al., 2010;

Upadhyay et al., 2015). Additionally, and as reported earlier, Naz et al. (2016) finds a positive association, and estimates an odds ratio of 1.30 between IAP and under-five mortality in India. Beyond child mortality, Balakrishnan et al. (2019) using data from Global Burden of Disease 2017, estimated that 1.2 million deaths in India (or 12.5% of the total deaths) were attributable to air pollution, including 0.7 million to ambient (outdoor) $PM_{2.5}$ and 0.5 million to IAP. Finally, Smith (2000) estimated that the annual health burden for India from IAP is 1.6-2.0 billion days of work lost (number of sick days due to the diseases caused by IAP) while Duflo et al. (2012) reports that a large portion of absence from schooling in rural areas of India is due to poor health.

Due to perceived health threats from polluting fuels, Indian authorities and non-governmental organizations (NGOs), have implemented policies and programs for reducing IAP. For example, subsidizing cleaner fuel technologies, distributing "improved cooking stoves", and convincing households to improve ventilation system within the household are common interventions. Among these policy strategies, the improved cook stove has become the most popular policy prescription for reducing IAP with the government of India implementing the second-largest program in the world to limit emission of smoke within households by distributing roughly 33 million biomass-based improved stoves in rural areas during 1984-2000 through its National Biomass Cookstoves Programme. However, these initiatives have received mixed reviews: while improved biomass stoves have reduced the time and effort that rural women put into collecting fuel per meal by half, it's effectiveness in reducing IAP and health benefits were far below the expectations. In fact, studies suggest that "improved" cooking stoves had a hazardous impact on health due to inefficient use (Hanbar and Karve, 2002; Kishore and Ramana, 2002).

2.2 Infant Mortality Trends in India

Figure 1 shows the trend in infant mortality by cooking fuel choices in India. Compared to the under-five mortality incidence that has leveled off at around 3.1% per year for households that use clean fuel for cooking, the under-five mortality rate remains more than twice as high for households using polluting fuel — although this rate has declined sharply by about 45%

over the past 25 years.[8] There is some variation in the mortality rate by age group: the neonatal mortality rate (defined as the probability of dying within the first 28 days of life) is the highest, followed by post-neonatal mortality (measured as the probability of dying between approximately the first month after birth and end of the first year of life) and then child mortality (assessed as the probability of dying between exact ages one and five). Decreasing trends are also observed for each age group, where the neonatal mortality rate declined from 4.4% in 1992 to 3.1% in 2016, post-neonatal mortality rate from 2.6% in 1992 to 1.1% in 2016, and child mortality rate from 1.5% in 1992 to 0.5% in 2016 for those using polluting fuels for cooking.[9]

3 Empirical Strategy

In this section, we first describe the empirical specification for the relationship between cooking fuel choice and child mortality. We then discuss the challenges in estimating the causal effect of cooking fuel choice on under-five mortality.

3.1 Indoor Air Pollution and Infant Mortality

To examine the causal effect of cooking fuel choice on mortality of children under five years of age, we specify the following relationship:

$$\begin{gathered} Y_{ihvdst} = \alpha + \beta D_{hvdst} + \boldsymbol{X}_{hvdst}\boldsymbol{\gamma} + \boldsymbol{M}_{jhvdst}\boldsymbol{\lambda} + \boldsymbol{W}_{ihvdst}\boldsymbol{\delta} + \\ + \mu_t + \eta_s + (\eta_s \times \mu_t) + \varepsilon_{ihvdst} \end{gathered} \tag{1}$$

where Y_{ihvdst} is one of the four binary variables for under-five mortality (neonatal, post-neonatal, child, and under-five) taking the value 1 if the death occurred during the considered age-periods, and 0 if the child survived during the age-period for child i, in household h, in

[8]The mortality rate (mortality incidence proportion, %) is calculated by the ratio (Number of child deaths/Total number of live births) for the trend analysis presented in Figure 1. In this paper, we will refer to mortality rate interchangeably with mortality incidence proportion.

[9]The mortality rates for each of the preceding three age-groups including neonatal, post-neonatal, and child add up to the under-five mortality rate. This is because (i) the three successive age groups constitute the first 5 years of life, and (ii) the mortality incidence proportions for different age groups have been calculated using a common denominator, total number of live births during the five-year window. The details about constructing the mortality measure has been provided in Section 4.

village v, in district d of state s, in survey year t. The key regressor is a binary variable for solid fuel use (D_{hvdst}) in household h, in village v, in district d of state s, in year t as defined above. The vectors $\boldsymbol{X}_{hvdst}$, $\boldsymbol{M}_{jhvdst}$, and $\boldsymbol{W}_{ihvdst}$ are respectively composed of household (h) characteristics including place of residence, household wealth index, number of household members, place where food is cooked and type of house, mother (j) characteristics including mother's age and mother's education, and child (i) characteristics including gender of the child and breastfeeding status. The error term, ε_{ihvdst}, captures the remaining unobserved, time-varying, and child-specific factors.

The state fixed effects, η_s, control for all permanent unobserved determinants of mortality across states, while the inclusion of year fixed effects for year of survey, μ_t, nonparametrically adjusts for national trends in under-five mortality, which is important in light of the time patterns observed in Figure 1. To control for possible unobserved spatial differences in cooking fuel at different periods, we interact the time fixed effect with the state fixed effect and include state-specific time trends, $\eta_s \times \mu_t$, to allow the unobserved time trend to vary across states.[10]

3.2 Identification

The key identification challenge is the potential endogeneity resulting from non-random use of polluting fuels. In the empirical literature on air pollution and its health consequences, it is commonly assumed that IAP affects mortality and other human health outcomes but not vice versa. In practice, IAP and choice of fuel types for cooking can be affected by mortality, morbidity, and other health outcomes. For example, Duflo et al. (2008b) document the potential impact of IAP on health, productivity, and ultimately long-term earnings. Noting that low-income households can only afford the cheaper fuel option which is frequently polluting and adversely affects health and earnings, we have a simultaneity issue that makes the choice of cooking fuel endogenous in Equation (1). We address this reverse causality from health outcomes to cooking fuel choice by estimating Equation (1) with instrumental

[10]Controlling for State×Time fixed effects allows us to estimate the effect of region-specific characteristics varying over time, which can be seen as regional (or neighborhood) differences such as culture, weather conditions, environmental features, and local-level policies or programs on cooking fuels.

variables (IVs).

A set of variables including speed of change in district forest cover over the period 2007–13 and household ownership status of agricultural land are tested as IVs both individually and combined, and the instruments are described in detail in the next section. Note that the variables measuring the relative change in tree cover over the given period are measured at the district level even though village-level information is available, for example, in the Census data. This is due to the random Primary Sampling Unit (PSU) point (or village/city block) displacement in the NFHS GPS data, which limits our ability to correctly match the PSUs with Census locations at the village-level.[11] In other words, we are unable to correctly match the NFHS dataset with Census and other datasets at sub-district (or *tehsil*) and village levels as the maximum displacement buffers for particular cluster points overlay with level 3 administrative (sub-district) boundaries. Figure 2 shows the displacement strategy of PSU points in NFHS-4 and the difficulty in correctly identifying the sub-districts and villages where the NFHS survey respondents reside. Although the PSU point displacement is random, it would affect our empirical analysis because we combine NFHS data with satellite and Census data by location.

Although we compute the speed of change in forest cover as a relative change in the percentage of forested area in the total geographical area using multiple years of satellite data, we have a single observation for each of the districts; thus, we are unable to use district fixed effects in Equation (1).

4 Data

Our empirical analysis is based on three datasets. The first data set (nearly 0.4 million observations) is nationally-representative data from India's National Family Health Survey (NFHS). The NFHS collects individual-level data on mortality incidence and other socio-economic characteristics for every member in the sample household. Additionally, it also

[11] According to the description of the NFHS GPS data provided by the DHS Program, the displacement is restricted so that the PSU points stay within the country, the NFHS survey region (state), and district area. Therefore, the displaced cluster's coordinates are located within the same country, state, and district areas as the undisplaced cluster. This random error can substantively affect analysis results, where analysis questions look at small geographic areas including sub-districts and villages/city blocks.

contains household-level information on wealth, housing, place of residence and agricultural land ownership status. Importantly, for our analysis, NFHS data includes information on the type of cooking fuel that household use, which allows us to approximate indoor air quality at the household level. To date, four rounds of the survey have been conducted since 1992–93.[12] Our analysis relies primarily on three rounds of this survey: NFHS-1 (1992–93), NFHS-2 (1998–99), and NFHS-4 (2015–16). We are unable to use the NFHS-3 (2005–06) in our empirical analysis due to the absence of district identifiers in the questionnaire of this particular round for confidentiality of HIV testing. A total of 1,003,880 ever-married women of reproductive ages between 15–49 years (317,250 from urban and 686,630 from rural areas) were included in the three surveys (NFHS-1, NFHS-2, and NFHS-4), that we analyzed in this paper. Ever-married women, aged less than 15, are excluded from the sample, and all the women interviewed in the survey were ever-married, of whom only 271 were aged less than 15 years. Our analysis is based on a pooled dataset of 421,709 singleton live-born children, of whom 22,268 died in the 5-years before the respective survey years.

Second, as a primary database on land use of the country, we use satellite data on forests from the Planning Commission of India. Third, from the 2011 Census of India, we also obtain land use information at the village and city block level. Specifically, we utilize the total surface area of the land of each geographical region and the land covered by forests, both measured in hectares.

4.1 Under-Five Mortality

Under-five mortality rates are an appealing measure of the effect of indoor air pollution for at least two reasons. First, children under five years tend to spend most of their time at home alongside their mothers, and since women are primarily responsible for cooking in India, under-five children tend to be exposed to indoor air pollution. Second, earlier years of life are especially vulnerable periods, and losses of life expectancy due to environmental exposure

[12] While the first three NFHS survey datasets cover all states of India, which includes more than 99% of India's population, the most recent NFHS data for the years 2015–16 (NFHS-4), adds all union territories for the first time. It is worth noting that we treat union territories as states. The NFHS-4 also provides vital estimates of most demographic and health indicators at the district level for all 640 districts in the country (as per the 2011 Census).

are likely to be large. Our primary outcome variable is under-five mortality. In addition, we consider three preceding age groups including neonatal, post-neonatal, and child mortality. Neonatal mortality is the number of deaths during the first 28 days of life (0–28 days), as a fraction of total number of live births; post-neonatal mortality is the number of deaths between one month and the first birthday (1–12 months), as a fraction of total number of live births; child mortality is the number of deaths between exact ages one and five (12–59 months), as a fraction of total number of live births.

4.2 Cooking Fuel Choice and Other Controls

The key explanatory variable in our analysis is the type of fuel used for cooking in the household, a proxy for indoor air pollution. Twelve types of cooking fuel are reported in the NFHS datasets, and we classify these fuels into two groups, clean and polluting, based on exposure to cooking smoke. The clean fuel includes electricity, liquid petroleum gas (LPG), natural gas and biogas, while polluting fuel includes kerosene, coal/lignite, charcoal, wood, straw/shrubs/grass, agricultural crop waste, and animal dung. Note that no household reported using more than one type of cooking fuel in the survey.

In addition to the main exposure variable, we collect information on several other determinants of under-five mortality. Place of residence (urban or rural), household wealth index (high wealth, middle wealth, or low wealth),[13] mother's education (secondary or higher, primary, or no education), type of house (pucca, semi-pucca, or kachha), and number of household members[14] are included as potential socio-economic factors (Wichmann and Voyi, 2006; Rinne et al., 2007; Tielsch et al., 2009; Bassani et al., 2010;

[13] The household wealth index was constructed using principal components analysis, with weights for the wealth index calculated by giving scores to the asset variables such as ownership of transport, durable goods, and facilities in the household. "Low wealth" referred to the bottom 40% of households, "middle wealth" referred to the middle 40% of households, and "high wealth" referred to the top 20% of households (Filmer and Pritchett, 2001).

[14] Number of household members refers to the total number of members living together in a household, which is not necessarily the same as family size. On average households in the survey have 7 members, but there are households with as many as 46 people. There is a strong positive correlation between household size and fuel choice, and the use of polluting fuels tends to increase as household size gets larger. Gas stove limits the volume of food that can be cooked because the size of the stove-top is small while wood burning furnaces can be built to accommodate larger utensils. The distribution of household size suggests that households with less than about 25 members are quite prevalent in the data while households with more than 25 members could be considered as outliers.

Epstein et al., 2013; Pandey and Lin, 2013; Ezeh et al., 2014; Naz et al., 2015, 2016).

Mother's age (<20, 20–29, 30–39, and 40–49 years) and gender of the child are also considered as potential confounders of the association between exposure to IAP and under-five mortality. Breastfeeding status of the mother (ever breastfed or never breastfed) and place where food is cooked (in the same room inside the house, in a separate room as kitchen inside the house, in a separate building, or outside)[15] are also factors that correlate with different levels of exposure to polluting fuels. No separate kitchen used for cooking inside the house has also been shown to be significantly associated with high exposure to IAP,[16] whereas breastfeeding has been shown to be a protective factor for under-five mortality, generally in the neonatal and post-natal periods.[17]

Thus, we can control for whether food is cooked inside the house, in a separate building, or outside using the data from NFHS-4 (2015–16) combined with an indicator for a separate kitchen inside the house.

4.3 Instruments for Cooking Fuel Choice

To account for the endogeneity of cooking fuel choice, we use forest cover to generate exogenous variation in the opportunity cost of cooking fuel choice. In the absence of data on prices of firewood and LPG, the main fuels for cooking in India, at district and/or village level we use forest cover as a proxy for the relative price (cost) of firewood.[18] We

[15]In the NFHS questionnaire, the question whether the household has separate room as kitchen captures only cooking inside the house and is not relevant to outdoor cooking. Another variable for indicating if the household cooks inside the house, in a separate building, or outdoors is only available in NFHS-4 (2015–16). Therefore, if we utilize this variable in our analysis, we are forced to use only the last round of NFHS survey. We separate cooking inside the house into two groups: in a separate room as kitchen inside the house or in the same room as they live inside the house based on variable indicating an existence of separate kitchen, i.e., question asking whether the household has a separate room used as kitchen in house. A separate room for cooking as compared to cooking inside is likely to be quite similar because of poor ventilation within houses, especially in rural areas, would lead to smoke permeating throughout if cooking with wood or coal.

[16]See, for example, Dasgupta et al. (2006); Khaliquzzaman et al. (2007); Khalequzzaman et al. (2010, 2011); Edwards and Langpap (2012); Gurley et al. (2013); Naz et al. (2015, 2016).

[17]See, for example, Cushing et al. (1998); Arifeen et al. (2001); Heinig (2001); Black et al. (2003); Wichmann and Voyi (2006); Ezeh et al. (2014).

[18]Kuo and Azam (2019) recently attempted to determine the drivers of household cooking fuel choice in India by estimating a panel multinomial logit regression with random effects based on two rounds of India Human Development Survey datasets. They show that access to paved road and peer effects significantly increase the probability of rural households to adopt clean fuel while distance to the nearest town is not an important driver of fuel choice in rural areas. In addition, Kuo and Azam (2019) find that the bargaining power or economic status of women in the household (proxied by education, financial independence and

expect that the speed of change in forest cover is exogeneous to child mortality.

The speed of change in forest cover is relevant and generates meaningful variation in cooking fuel choice through several channels. First, wood is the most widely-used fuel for cooking in India. Figure 3 shows that one-half of the Indian households covered in four rounds of the NFHS rely on wood as a fuel for cooking. The speed of change in forest cover generates variations in access to or availability of firewood (polluting) for cooking, and households living in villages with forest use firewood twice as much as households in villages without forest (Pinto et al., 1985). Figure 4 illustrates India's district-wise forest cover as in 2011 by utilizing satellite-based information from the Planning Commission of India. The share of households using solid fuels for cooking in three of the largest five forest cover states (88% in Odisha, 84% in Chhattisgarh, and 81% in Madhya Pradesh) is substantially larger than the country average, 76%, suggesting that location of forests affects cooking fuel choice. Also note that the correlation coefficient between the speed of change in forest cover and the 2011 level of forest cover is 0.9984 (SE: 0.0024, p-value: 0.00) at the district level. Since the speed of change in forest cover is positively correlated with the level of forest cover, our argument here is also applicable for regions with high speed of growth in forest cover. Furthermore, under-five mortality rates in these three states (6.0% in Odisha, 5.8% in Chhattisgarh, and 6.6% in Madhya Pradesh) are persistently higher than the country average, 5.3%. This geographic variable hence induces plausibly exogenous variation in cooking fuel choice that is not correlated with the unobserved, time-varying, and child-specific shocks to under-five mortality.

We obtain district-level satellite data on forest cover from the Planning Commission of India (reformed as the National Institution for Transforming India–NITI Aayog in 2015) for three years, including 2007, 2011, and 2013. The baseline regressions use the speed of change or relative change in forest cover, where forest cover is defined by forested area as a percentage of total geographical area, based on data from the NITI Aayog to account for the spatial and temporal variation in forest cover. A change in forest cover over the periods 2011 and 2013 minus a change in forest cover over the periods 2007 and 2011 measures the

freedom) and price of LPG are critical for urban households to make a decision about adopting clean fuel. However, the determinants of fuel choice have to affect the child mortality only through cooking fuel choice in order to be valid instruments.

speed of change or relative change in forest cover over the three years. In the Planning Commission dataset, forest cover refers to all lands more than one hectare in area, with a tree canopy density of more than 10 percent irrespective of ownership and legal status. It also includes orchards, bamboo, and palm. The satellite-based tree cover has been classified, based on tree canopy density, into four categories including very dense forest, moderately dense forest, open forest, and scrub, and we consider the first three of these forest types in our analysis excluding the scrub.

An alternative measure of forest cover is available from the 2011 Indian Census which provides village-level data on land covered by forests (in hectares). We define forest cover as per-capita forest area (ha/person, unreported) and percentage of total geographical area of the village under forest (% of land area). The village-level data on population and the geographical area of the village also come from the 2011 Census of India. Because areas inhabited by tribal population and inaccessible hilly geographic areas present a problem in nationwide ground-level census of trees in India (Foster and Rosenzweig, 2003), we prefer the satellite-based data as our primary measure of forest cover and use the census-based measure as a robustness check. The bivariate correlation of satellite-based forest cover with census-based forest cover is 0.62. This shows that the Indian census- and satellite-based tree cover data are indeed different but quite comparable.

Table 1 presents summary statistics on cooking fuels, infant mortality, and other demographic indicators used in the regression analysis. As can be seen, the data suggests that under-five mortality rate in India during the period analyzed was 5.3%, and infant mortality rate increases as age of the child decreases. A majority (76.0%) of the households use polluting fuels, while the remaining households use clean (electricity, LPG, and biogas) fuels for cooking. Three-fourth of the children included in our analysis are from rural areas. Overall across rural and urban areas 67.9% of the mothers with children aged under five are in the 20-29 years old age bracket. In terms of other socio-economic characteristics, including household wealth, mother's education, gender of child, location where food is cooked, and type of house, the individuals included in the analysis are evenly distributed.

Tables 2 and 3 provides the mean and standard deviation of the four outcome variables (infant mortality for different age-groups) and key explanatory variable (type of cooking fuel) by geographic region, age and gender of the household head, along with the associated number of observations. Evidently, infant mortality rate and fuel choices significantly vary across regions throughout the country (Table 2). By contrast, infant mortality and fuel choices are relatively stable across different age groups (Panel A of Table 3) and gender (Panel B of Table 3) of the household head.

5 Results

In this section, we first present the estimated average marginal effects[19] of cooking fuel choice on child mortality using a multivariate probit and the IV (2SLS) regressions. We then discuss the implications of our baseline results and present a set of robustness tests. We begin with the probit model results to create a comparable benchmark against the existing literature.

5.1 Probit Estimates

Table 4 presents the results of estimating Equation (1) as a pooled probit model for under-five mortality under three different specifications with more control variables added successively. The average marginal effect (AME) of the key regressor, use of polluting fuel for cooking, ranges from 2.3 to 0.8 percentage points in the three regressions. The basic model shown in Column (1) includes year and state fixed effects and is estimated using NFHS-1, NFHS-2,

[19]Marginal effects are computed using two methods: average marginal effects (AME) and marginal effects at the means (MEM). MEM is calculated by setting the values of all covariates to their means within the sample. On the other hand, to obtain the AME, the marginal effect is first calculated for each individual with their observed levels of covariates, and these values are then averaged across all individuals. Since our independent variables, except for the number of household members, including our key regressor, fuel choice, are binary variables, the average marginal effects measure *discrete change* or how the predicted probabilities (infant mortality) change as the binary independent variables change from 0 to 1. For probit regression, the average marginal effect of $\mathbf{x}_k = (x_{1k} \cdots x_{ik} \cdots x_{Nk})'_{(N\times 1)}$ on $\boldsymbol{y} = (y_1 \cdots y_i \cdots y_N)'_{(N\times 1)}$ is calculated by

$$AME = \frac{1}{N}\sum_{i=1}^{N}\left[\Phi\Big(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}|x_{ik}=1\Big) - \Phi\Big(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}|x_{ik}=0\Big)\right]$$

where Φ is the probability density function for a standardized normal variable, and $\boldsymbol{x}_i' = (x_{i1} \cdots x_{ik} \cdots x_{iK})_{(1\times K)}$ is a vector of explanatory variables. Intuitively, for example, the AME of fuel choice demonstrates that a change of *polluting fuel for cooking* from 0 to 1 changes the probability that the *under-five mortality* takes the value of 1 by how many percentage points. There are several ways to compute the standard errors for the AMEs of regressors. The standard errors of the AMEs in this paper have been computed using the delta method, which is a semi-parametric method for deriving the variance of a function of asymptotically normal random variables with known variance.

and NFHS-4, while the probit models, shown in Columns (2) and (3), are estimated using only NFHS-4 because a variable capturing an actual place where food was cooked is only available in the last round of survey. Since the coefficient estimates and calculated marginal effects of polluting fuel use are consistently greater than zero and statistically significant at 1 percent level for each specification, we conclude that the use of polluting fuels for cooking is associated with the mortality risk amongst children aged under-five in India. We consider the last regression as our preferred or primary specification because the inclusion of state-by-year fixed effects controls for time-varying spatial factors such as state attributes (e.g., characteristics of state magistrate, whether there is any government program regarding the child health service in the state, and access to medical facilities) and local characteristics (e.g., distance from urban areas and large cities, percentage of districts, tehsils, or villages with paved roads, outdoor air quality, and quality of soil and water resources) that could affect both under-five mortality and fuel choice.

To examine the effect of cooking fuel choice on infant mortality in more detail, we consider three alternative age groups: neonatal, post-neonatal, and child. Table 5 presents the results for child mortality. The average marginal effect of IAP on child mortality decreases significantly from 0.8 to 0.2 percentage points as well as the associations of other confounders change dramatically in magnitude. This major decrease is quite intuitive because the most vulnerable period (or the first year of life) has been excluded from the first five years of life. In other words, childhood between ages one and five is a less risky period compared to neonatal and post-neonatal periods which are included in under-five years of age. The results for the post-neonatal mortality are presented in Table 6. The average marginal effect of IAP on post-neonatal mortality is estimated at 0.1 percentage point; however, it is not statistically significant.

Table 7 shows the results for neonatal mortality. Compared to other two relatively older age groups, average marginal effect of polluting fuel choice on neonatal mortality is estimated at 0.6 percentage point, the largest estimate among these three alternative age groups. One would expect that the youngest age group should have the largest coefficient estimate since the neonatal period is the most vulnerable time for a child's survival. Overall our results suggest that the harmful effect of IAP on infant mortality increases for the youngest children, which is consistent with the existing child's age-risk of dying (or -child's vulnerability) argument. A comparison between baseline results in Table 4 and those under the three alternative outcomes in Tables 5–7 suggests that the key results are robust to the range of plausible age differences of child mortality from the literature. An important implication of this finding is that the harmful effect of IAP can be reduced by improving the care for infants to increase the immunity.

The average marginal effects of the other variables are all intuitively signed and are consistent with the infant mortality literature. The risk of mortality in mothers who had never breastfed is the highest compared to other confounders, which is in line with previous findings (Cushing et al., 1998; Arifeen et al., 2001; Heinig, 2001; Black et al., 2003; Wichmann and Voyi, 2006; Ezeh et al., 2014). While infant mortality is positive and significant for teenage mothers, older mothers (in age groups 20-29 and 30-39) have a lower risk of under-five child mortality. Our results also show that mother's education is inversely related to under-five mortality. Infant mortality is also higher in households of middle- and low-wealth compared to the high-wealth ones, households with no separate kitchen inside the house, and households that live in semi-pucca and kaccha (makeshift and temporary) houses. Cooking outside is essentially the same as cooking in the living room in terms of their association with infant mortality (Column (3) of Tables 4–7), possibly due to poor ambient air quality. We also estimate specifications with *district* and *district* $\times$ *time* fixed effects and obtain qualitatively identical results.

In Table 8, we compare our results from a nonlinear model with those from Naz et al. (2016) which uses a multivariate logistic regression and data from NFHS-1 (1992–93), NFHS-2 (1998-99) and NFHS-3 (2005–06) to estimate an association between use of polluting fuel for cooking and infant mortality. The results (or odds ratio) of Naz et al. (2016) are reported in Column (1), while our replication results and corresponding calculated average marginal effects are shown in Column (2). Since our analysis utilizes the most recent round of NFHS, or NFHS-4 (2015–16), we also estimate simple logistic regression with the same specification as Naz et al. (2016) using NFHS-4 data. Columns (3) and (4) present the estimated odds ratios and corresponding marginal effects using only NFHS-4 (2015–16) and a complete sample between 1992–2016 (NFHS-1-4), respectively. Compared with our primary specification (Column (5) of Table 8) which includes additional controls for the location where food is cooked (inside/outside/separate room of the house) and a set of fixed effects, the replicated (or Naz et al. (2016)) average marginal effects of polluting fuel use on infant mortality are almost always higher.

5.2 Linear IV Estimates

We address the endogeneity of cooking fuel choice using IV strategy. We explore the speed of change in forest cover and agricultural land ownership respectively as a region and household-specific characteristics, which create exogenous variations in fuel choice of the households and serve as IVs for our endogenous variable.[20]

We first present evidence on how speed of change in forest cover and agricultural land ownership relate to household's choice of fuel types used for cooking. The relationships are estimated using linear model, where the dependent variable is a binary variable whether fuel choice. The correlation coefficients of speed of change in forest cover and agricultural land ownership with mean fractions of polluting fuel use for cooking are 0.0824 (SE: 0.0430, p-value: 0.06) and 0.5816 (SE: 0.0332, p-value: 0.00) at the district-level, respectively.

Column (1) of Table 9 reports the first-stage results when the indicator variable for household's agricultural land ownership is used as an IV. Agricultural land ownership is a dummy variable and takes the value of 1 if household owns land for agricultural purposes in

[20]The bivariate correlations of under-five mortality with agricultural land ownership and speed of change in forest cover are 0.0038 (SE: 0.0020, p-value: 0.06) and -0.0110 (SE: 0.0021, p-value: 0.00), respectively. One may argue that infant mortality is negatively associated with ownership of agricultural land through an income channel considering that agricultural production is a source of household revenue. The negative relationship between infant mortality and household wealth is illustrated in Figure 5a. However, Figure 5b shows that agricultural land ownership status is negatively associated with the household wealth. This means that variation in agricultural land ownership is not necessarily a proxy for variation in household wealth in India. This observation is consistent with the fact that there are many small farm households in India. Hence, it suggests that the household's status of agricultural land ownership does not necessarily indicate that a family is wealthy, supporting the idea that agricultural land ownership status at least does not affect the household fuel choice through the income channel. Note that we use household wealth (stock) as a proxy of household's income (flow) given that the DHS data does not include actual earnings of the household. Our argument here holds if household wealth represents its income. We find (unreported) a negative and statistically significant relationship between agricultural land ownership and principal components of household wealth index which require households to have flows of income to operate them (or with variable cost) such as ownership of refrigerator, television, washing machine, electric fan, air conditioner or cooler, and computer, conditional on a set of time and spatial fixed effects. Thus, we can assume that wealth and income are correlated. Additionally, some of our other controls, such as mother's education and the number of household members, potentially capture the household income.

a given year. This variable has a positive and statistically significant impact on cooking fuel choice. Agricultural households are likely to consume their own agricultural crop waste and animal dung as cooking fuel which are classified as polluting. This confirms that agricultural land ownership generates plausible variation in fuel choice. Columns (2)–(5) of Table 9 present the estimates from the IV (2SLS) regressions for four different age groups. The coefficient estimates on polluting fuel for cooking for under-five and neonatal mortality are positive and statistically significant, ranging from 0.037 to 0.050.

Column (1) of Table 10 presents the first-stage results when speed of change in forest cover and an indicator variable for household's agricultural land ownership are jointly included as IVs. The joint F-statistic on the excluded instruments is large enough to suggest that these two IVs provide plausible variations in fuel choice that we can leverage to identify a causal effect of fuel choice on infant mortality.[21] Columns (2)–(5) in Panel A present the estimates from the IV (2SLS) regressions for four different age groups. The coefficient estimates for polluting fuel for cooking for under-five and neonatal mortality are positive and statistically significant at 5 percent level, ranging from 0.034 to 0.049. In other words, a family relying on polluting fuel for cooking has a 4.9 and 3.4 percent higher probability of experiencing child mortality in the first five years and within the first 28 days of life, respectively. Heteroskedasticity-consistent standard errors were clustered at the district level instead of PSU level, given the utilization of district-level speed of change in forest cover as one of the instruments. The Hansen's J-statistic suggests that the excluded IVs are exogenous.

Panel B of Table 10 reports the coefficient estimates from OLS model to show how employing IV strategy affects our simple model estimates. Compared to the coefficient estimates from a simple OLS model (Columns (2) and (5) in Panel B), IV regression coefficients suggest that addressing endogeneity in the fuel choice-infant mortality

[21] Although there is theoretically no concern about the "relatively large" value of F-statistic on excluded IVs, in practice, one may be concerned about it. The "large" value of F-stat on IVs is possibly due to (i) a large sample, and (ii) "perfect" multicollinearity between instruments and an endogenous regressor. The latter would indicate that the instruments are not exogenous. This would be the case if R^2 of the first-stage regression is "too large" and household fuel choice is perfectly correlated with the speed of change in forest cover and agricultural land ownership. The R^2 of 0.54 shown in Column (1) of Table 10 indicates that our endogenous regressor not perfectly correlated with the instruments. Hence, we consider that the value of F-statistic reflects our sample size.

relationship leads to about six-fold increase in the estimates of the causal impact on under-five and neonatal mortality. The causal effect of polluting fuel use on child mortality essentially becomes zero.

5.3 Robustness Checks

To assess the robustness of our findings, we re-estimate the causal effect of cooking fuel choice on infant mortality using (two-step) IV probit regressions as an alternative to the IV regression. We find that IV probit provides exactly the same conclusion as the IV (2SLS) regression, verifying that the results are robust to an alternative estimation approach.

Panel A of Table A.1 presents the parameter estimates derived from the IV probit regression for under-five, child, post-neonatal, and neonatal mortality (with the same specification as used in Panel A of Table 10 where both relative change in forest cover over time and agricultural land ownership are used as IVs). It shows that using dirty fuels instead of clean fuels causes under-five and neonatal mortality, and the corresponding coefficient estimates on polluting fuel for cooking ranges from 0.556 to 0.569. The effect of employing an IV strategy on the magnitude of the causal effect of polluting fuel use on under-five and neonatal mortality is also similar to IV (2SLS) regression. In particular, both estimates of the causal effect of under-five and neonatal mortality also increase approximately six-fold once the endogeneity in the fuel choice-infant mortality relationship is addressed in comparison with the coefficient estimates from a pooled probit model (Columns (2) and (5) in Panel B of Table A.1). The causal impact of polluting fuel use on child mortality also becomes zero. Note that the pooled probit models in Columns (2)-(5) of Panel B of Table A.1 are exactly the same as those in Column (3) of Tables 4-7, respectively.

Second, the National Biomass Cookstoves Initiative (NBCI) was launched by the Indian government to enhance the use of improved biomass cookstoves in 2009. The pilot projects distributed 12,000 improved cookstoves to households in the states of Jammu and Kashmir, Uttar Pradesh, Bihar, Madhya Pradesh, Jharkhand, Chhattisgarh, Karnataka, and Odisha.[22] Hence, we additionally control for states where improved cookstoves program has

[22] The government of India had also initiated the National Programme on Improved Chulha (NPIC) in

been implemented by adding a dummy variable which indicates states where there is NBCI program. Notice that we omit the state fixed effect for one of the NBCI states to resolve the collinearity problem. It is important to note that we did not control for states with another government program, National Programme on Improved Chulha (NPIC), since it already became a nationally disseminated program. Table A.2 reports results from the first and second-stage regressions of IV (2SLS) regression where a dummy variable is added to our preferred specifications in Table 10. The first stage regression results suggest that the effect of the NBCI program on household fuel choice is not statistically significant, which is consistent with the existing findings from the literature including Hanna et al. (2016). The results obtained with the inclusion of the dummy variable for NBCI implementation are qualitatively identical to the IV (2SLS) regression results.

Third, we disaggregate our key regressor by ranking fuel types from 1 (the cleanest fuel) to 10 (the dirtiest fuel) based on their cleanliness or the energy ladder concept (Goldemberg, 2000). The assigned values to different types of fuels used for cooking are: 1 = electricity, 2 = LPG/natural gas, 3 = biogas, 4 = kerosene, 5 = coal/lignite, 6 = charcoal, 7 = wood, 8 = straw/shrubs/grass, 9 = agricultural crop, and 10 = animal dung. Table A.3 shows that if dirtiness level of cooking fuel increases by 1 unit, the probability of under-five and neonatal mortality will rise by 0.9 and 0.6 percent, respectively. In other words, the probability of experiencing child mortality within five years and 28 days of birth increases respectively by 0.9 and 0.6 percent if a household switches to a fuel type that is dirtier by one level along the energy ladder. Notice that the key regressor here, dirtiness level of cooking fuels, is a categorical variable. Our results here remain remarkably similar to our baseline results that use the cooking fuel choice as a binary variable.

Finally, we leverage satellite- and census-based data on forest cover (% of geographical area) in 2011 to test whether we can still identify a positive impact of polluting fuel use on under-five and neonatal mortality incidences. Using data on 2011 satellite-based forest cover and tree cover from the 2011 Indian Census as alternates to a satellite-based speed of change in forest cover, we find that the results are also exceptionally robust to the utilization of

1984 to provide efficient cooking stoves to rural areas in an attempt to limit air pollution. NPIC became a nation-wide program in 1986 and was implemented until 2000. Since this program had universal coverage throughout the country, we cannot use this program for a robustness check.

either satellite- or census-based tree cover for a single year as one of the IVs for household fuel choice (Tables A.4 and A.5, respectively).

6 Conclusion

Almost half of the global population continues to rely on solid fuels for cooking, and it constitutes largest source of indoor air pollution. In 2015, 64% of the Indian population used different types of solid fuels for cooking including wood, dung and coal, second after Sub-Saharan Africa. Each year, diseases attributed to indoor air pollution (IAP) kill 1.2 million people, including 100,000 children in India. Leveraging a unique and large-scale household survey data from 1992 to 2016 and geospatial information of forest cover in India, we find that the use of solid fuels for cooking increases under-five mortality and that our results are robust to a variety of empirical specifications.

Our analysis presents two important departures from the existing literature. First, we utilize nationally-representative demographic survey data instead of focusing on RCTs conducted in a particular region of the country as commonly analyzed in the literature (Diaz et al., 2007; Smith-Sivertsen et al., 2009; Hanna et al., 2016). Our analyses based on simple probit regressions lead to a 0.6 percentage points decrease in the estimates of the marginal impact of cooking fuel on infant mortality relative to the extant literature. This suggests that the literature has tended to overestimate the association between IAP and under-five mortality by approximately 152,000 deaths per year nationally as compared to our estimates. Our non-IV estimation departs from the existing literature in terms of additional controls and a more recent sample which points to the importance of including a full set of controls.

Second, ours is the first empirical analysis to address the endogeneity in cooking fuel choice when quantifying the causal effect of cooking fuels on infant mortality. The speed of change in forest cover and agricultural land ownership status in India provide plausibly exogenous variation in cooking fuels for causal identification. The IV (2SLS) analysis based on the speed of change in forest cover and agricultural land ownership shows that a household using solid fuel for cooking has a 3.4 and 4.9 percent higher probability of experiencing child mortality within 28 days and five years of birth, respectively. However, we find no causal impact of fuel choice on post-neonatal and child mortality.

We conclude with some caveats and directions for future research. First, our analysis is based on an indirect indicator of IAP, i.e., type of cooking fuels, to estimate the effect of IAP on under-five mortality due to the lack of data availability. Using direct measures of IAP (CO and PM emissions in homes) recorded by 24-hour carbon monoxide readings might provide more accurate estimates. Although cooking is the main source of IAP, it is not the only source of CO emission inside the house that poses risks to children's health. The WHO guidelines for household fuel combustion (WHO, 2014) classify kerosene as a polluting fuel and discourage its use as a household fuel. Nevertheless, kerosene is still used for lighting by around one billion people who lack access to electricity. Kerosene lamps are often the only means of lighting houses at night. Use of kerosene not only pollutes the air inside the house but also increases the risks for fires, burns and CO poisoning. Therefore, we might have underestimated the effect of IAP on infant mortality due to the absence of a direct measure of IAP and indirect measures for other sources of household air pollution.

Second, we focus on the causal impact of IAP on infant mortality. It is well understood that IAP affects not only infant mortality but also other socio-economic and health outcomes. Hence, future research could empirically examine the impact of cooking fuels on productivity of children and adults, school attendance, labor market participation, all of which could have important implications on the broader economy and contribute to the economic literature of indoor air quality or fuel choice.

References

Apte, Komalkirti, and Sundeep Salvi. 2016. "Household Air Pollution and Its Effects on Health." *F1000Research*, 5.

Arifeen, Shams, Robert E. Black, Gretchen Antelman, Abdullah Baqui, Laura Caulfield, and Stan Becker. 2001. "Exclusive Breastfeeding Reduces Acute Respiratory Infection and Diarrhea Deaths among Infants in Dhaka Slums." *Pediatrics*, 108(4): e67–e67.

Balakrishnan, Kalpana, Sagnik Dey, Tarun Gupta, R.S. Dhaliwal, Michael Brauer, Aaron J. Cohen, Jeffrey D. Stanaway, Gufran Beig, Tushar K. Joshi, Ashutosh N. Aggarwal et al. 2019. "The Impact of Air Pollution on Deaths, Disease Burden, and Life Expectancy across the States of India: The Global Burden of Disease Study 2017." *The Lancet Planetary Health*, 3(1): e26–e39.

Barron, Manuel, and Maximo Torero. 2017. "Household Electrification and Indoor Air Pollution." *Journal of Environmental Economics and Management*, 86: 81–92.

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): 491–499.

Bharadwaj, Prashant, Matthew Gibson, Joshua Graff Zivin, and Christopher Neilson. 2017. "Gray Matters: Fetal Pollution Exposure and Human Capital Formation." *Journal of the Association of Environmental and Resource Economists*, 4(2): 505–542.

Black, Robert E., Saul S. Morris, and Jennifer Bryce. 2003. "Where and Why are 10 Million Children Dying Every Year?" *The Lancet*, 361(9376): 2226–2234.

Bonjour, Sophie, Annette Prüss-Üstün, and Eva Rehfuess. 2007. *Indoor Air Pollution: National Burden of Disease Estimates*. World Health Organization.

Bruce, Nigel, John McCracken, Rachel Albalak, Morten Scheid, Kirk R. Smith, Victorina Lopez, and Chris West. 2004. "Impact of Improved Stoves, House Construction and Child Location on Levels of Indoor Air Pollution Exposure in Young Guatemalan Children." *Journal of Exposure Science and Environmental Epidemiology*, 14(S1): S26–S33.

Chang, T., Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2016. “Particulate Pollution and the Productivity of Pear Packers.” *American Economic Journal: Economic Policy*, 8(3): 141–169.

Chang, T., Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2019. “The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China.” *American Economic Journal: Applied Economics*, 11(1): 151–172.

Cushing, Alice H., Jonathan M. Samet, William E. Lambert, Betty J. Skipper, William C. Hunt, Stephen A. Young, and Leroy C. McLaren. 1998. “Breastfeeding Reduces Risk of Respiratory Illness in Infants.” *American Journal of Epidemiology*, 147(9): 863–870.

Dasgupta, Susmita, Mainul Huq, M. Khaliquzzaman, Kiran Pandey, and David R. Wheeler. 2006. “Indoor Air Quality for Poor Families: New Evidence from Bangladesh.” *Indoor Air*, 16(6): 426–444.

Diaz, Esperanza, Tone Smith-Sivertsen, Dan Pope, Rolv T. Lie, Anaite Diaz, John McCracken, Byron Arana, Kirk R. Smith, and Nigel Bruce. 2007. “Eye Discomfort, Headache and Back Pain among Mayan Guatemalan Women Taking Part in a Randomised Stove Intervention Trial.” *Journal of Epidemiology & Community Health*, 61(1): 74–79.

Centers for Disease Control and Prevention. 2006. *Principles of Epidemiology in Public Health Practice: An Introduction to Applied Epidemiology and Biostatistics*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), Office of Workforce and Career Development.

Duflo, Esther, Michael Greenstone, and Rema Hanna. 2008a. “Cooking Stoves, Indoor Air Pollution and Respiratory Health in Rural Orissa.” *Economic and Political Weekly*, 71–76.

Duflo, Esther, Michael Greenstone, and Rema Hanna. 2008b. “Indoor Air Pollution, Health and Economic Well-being.” *SAPI EN. S. Surveys and Perspectives Integrating*

Environment and Society(1.1).

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review*, 102(4): 1241–1278.

Edwards, John H.Y., and Christian Langpap. 2012. "Fuel Choice, Indoor Air Pollution and Children's Health." *Environment and Development Economics*, 17(4): 379–406.

Epstein, M.B., M.N. Bates, N.K. Arora, K. Balakrishnan, D.W. Jack, and K.R. Smith. 2013. "Household Fuels, Low Birth Weight, and Neonatal Death in India: The Separate Impacts of Biomass, Kerosene, and Coal." *International Journal of Hygiene and Environmental Health*, 216(5): 523–532.

Ezeh, Osita Kingsley, Kingsley Emwinyore Agho, Michael John Dibley, John Joseph Hall, and Andrew Nicolas Page. 2014. "The Effect of Solid Fuel Use on Childhood Mortality in Nigeria: Evidence from the 2013 Cross-sectional Household Survey." *Environmental Health*, 13(1): 113–122.

Filmer, Deon, and Lant H. Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data—or Tears: An Application to Educational Enrollments in States of India." *Demography*, 38(1): 115–132.

Foster, Andrew D., and Mark R. Rosenzweig. 2003. "Economic Growth and the Rise of Forests." *The Quarterly Journal of Economics*, 118(2): 601–637.

Goldemberg, José. 2000. "Rural Energy in Developing Countries." In *Energy and the Challenge of Sustainability.* United Nations Development Programme, 367–389.

Graff Zivin, Joshua, and Matthew Neidell. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review*, 102(7): 3652–3673.

Graff Zivin, Joshua, and Matthew Neidell. 2013. "Environment, Health, and Human Capital." *Journal of Economic Literature*, 51(3): 689–730.

Graff Zivin, Joshua, and Matthew Neidell. 2018. "Air Pollution's Hidden Impacts." *Science*, 359(6371): 39–40.

Greenland, Sander, and Kenneth J. Rothman. 2008. "Measures of Occurrence." In *Modern Epidemiology.* Lippincott Williams & Wilkins, 32–50.

Gurley, Emily S., Nusrat Homaira, Henrik Salje, Pavani K. Ram, Rashidul Haque, William Petri, Joseph Bresee, William J. Moss, Patrick Breysse, Stephen P. Luby, and Eduardo Azziz-Baumgartner. 2013. "Indoor Exposure to Particulate Matter and the Incidence of Acute Lower Respiratory Infections among Children: A Birth Cohort Study in Urban Bangladesh." *Indoor Air*, 23(5): 379–386.

Hanbar, R.D., and Priyadarshini Karve. 2002. "National Programme on Improved Chulha (NPIC) of the Government of India: An Overview." *Energy for Sustainable Development*, 6(2): 49–55.

Hanna, Rema, Esther Duflo, and Michael Greenstone. 2016. "Up in Smoke: The Influence of Household Behavior on the Long-run Impact of Improved Cooking Stoves." *American Economic Journal: Economic Policy*, 8(1): 80–114.

Hanna, Rema, and Paulina Oliva. 2015. "The Effect of Pollution on Labor Supply: Evidence from a Natural Experiment in Mexico City." *Journal of Public Economics*, 122: 68–79.

Heinig, M. Jane. 2001. "Host Defense Benefits of Breastfeeding for the Infant: Effect of Breastfeeding Duration and Exclusivity." *Pediatric Clinics of North America*, 48(1): 105–123.

Imelda. 2018. "Indoor Air Pollution and Infant Mortality: A New Approach." In *AEA Papers and Proceedings*. 108: 416–421, American Economic Association.

Imelda. 2019. "Cooking that Kills: Cleaner Energy, Indoor Air Pollution, and Health." *mimeo*. Department of Economics, Universidad Carlos III de Madrid.

Isen, Adam, Maya Rossin-Slater, and W. Reed Walker. 2017. "Every Breath You Take—Every Dollar You'll Make: The Long-term Consequences of the Clean Air Act of 1970." *Journal of Political Economy*, 125(3): 848–902.

Khalequzzaman, M., Michihiro Kamijima, Kiyoshi Sakai, Takeshi Ebara, Bilqis Amin Hoque, and Tamie Nakajima. 2011. "Indoor Air Pollution and Health of Children in Biomass Fuel-Using Households of Bangladesh: Comparison between Urban and Rural Areas." *Environmental Health and Preventive Medicine*, 16(6): 375–383.

Khalequzzaman, M., Michihiro Kamijima, Kiyoshi Sakai, Bilqis Amin Hoque, and Tamie Nakajima. 2010. "Indoor Air Pollution and the Health of Children in Biomass-and Fossil-Fuel Users of Bangladesh: Situation in Two Different Seasons." *Environmental Health and Preventive Medicine*, 15(4): 236–243.

Khaliquzzaman, M., Michihiro Kamijima, K. Sakai, N.A. Chowdhury, N. Hamajima, and T. Nakajima. 2007. "Indoor Air Pollution and Its Impact on Children Under Five Years Old in Bangladesh." *Indoor Air*, 17(4): 297–304.

Kishore, V.V.N., and P.V. Ramana. 2002. "Improved Cookstoves in Rural India: How Improved Are They? A Critique of the Perceived Benefits from the National Programme on Improved Chulhas (NPIC)." *Energy*, 27(1): 47–63.

Kuo, Ying-Min, and Mehtabul Azam. 2019. "Household Cooking Fuel Choice in India, 2004-2012: A Panel Multinomial Analysis." *IZA Discussion Paper* No. 12682.

Martin, William J., John W. Hollingsworth, and Veerabhadran Ramanathan. 2014. "Household Air Pollution from Cookstoves: Impacts on Health and Climate." In *Global Climate Change and Public Health*. Springer, 237–255.

Menon, Premlata. 1988. "Indoor Spatial Monitoring of Combustion Generated Pollutants (TSP, CO, and BaP) by Indian Cookstoves." *mimeo*. Department of Meteorology, University of Hawaii.

Naz, Sabrina, Andrew Page, and Kingsley Agho. 2015. "Household Air Pollution and Under-Five Mortality in Bangladesh (2004–2011)." *International Journal of Environmental Research and Public Health*, 12(10): 12847–12862.

Naz, Sabrina, Andrew Page, and Kingsley Emwinyore Agho. 2016. "Household Air Pollution and Under-Five Mortality in India (1992–2006)." *Environmental Health*, 15(1): 54–64.

Pandey, Shanta, and Yuan Lin. 2013. "Adjusted Effects of Domestic Violence, Tobacco Use, and Indoor Air Pollution from Use of Solid Fuel on Child Mortality." *Maternal and Child Health Journal*, 17(8): 1499–1507.

Pinto, A.G., P. Khanna, and C.N. Pandey. 1985. "Gujarat Wood Balance Study 1984."

mimeo. Gujarat State Forest Department, Vadodara.

Pitt, Mark M., Mark R. Rosenzweig, and M. Nazmul Hassan. 2006. "Sharing the Burden of Disease: Gender, the Household Division of Labor and the Health Effects of Indoor Air Pollution in Bangladesh and India." In *Stanford Institute for Theoretical Economics Summer Workshop*. 202.

Rehfuess, Eva A., N.G. Bruce, and K.R. Smith. 2011. "Solid Fuel Use: Health Effect." *Encyclopedia of Environmental Health*, 5: 150–161.

Rinne, Seppo T., Edgar J. Rodas, Mikael L. Rinne, Joshua M. Simpson, and Larry T. Glickman. 2007. "Use of Biomass Fuel is Associated with Infant Mortality and Child Health in Trend Analysis." *The American Journal of Tropical Medicine and Hygiene*, 76(3): 585–591.

Saksena, S., R. Prasad, R.C. Pal, and V. Joshi. 1992. "Patterns of Daily Exposure to TSP and CO in the Garhwal Himalaya." *Atmospheric Environment. Part A. General Topics*, 26(11): 2125–2134.

Schindler, Mirjam, Geoffrey Caruso, and Pierre Picard. 2017. "Equilibrium and First-Best City with Endogenous Exposure to Local Air Pollution from Traffic." *Regional Science and Urban Economics*, 62: 12–23.

Silwal, Ani Rudra, and Andy McKay. 2015. "The Impact of Cooking with Firewood on Respiratory Health: Evidence from Indonesia." *The Journal of Development Studies*, 51(12): 1619–1633.

Smith, Kirk R. 2000. "National Burden of Disease in India from Indoor Air Pollution." *Proceedings of the National Academy of Sciences*, 97(24): 13286–13293.

Smith-Sivertsen, Tone, Esperanza Diaz, Dan Pope, Rolv T. Lie, Anaite Diaz, John McCracken, Per Bakke, Byron Arana, Kirk R. Smith, and Nigel Bruce. 2009. "Effect of Reducing Indoor Air Pollution on Women's Respiratory Symptoms and Lung Function: The RESPIRE Randomized Trial, Guatemala." *American Journal of Epidemiology*, 170(2): 211–220.

Tielsch, James M., Joanne Katz, Ravilla D. Thulasiraj, Christian L. Coles,

S. Sheeladevi, Elizabeth L. Yanik, and Lakshmi Rahmathullah. 2009. "Exposure to Indoor Biomass Fuel and Tobacco Smoke and Risk of Adverse Reproductive Outcomes, Mortality, Respiratory Morbidity and Growth among Newborn Infants in South India." *International Journal of Epidemiology*, 38(5): 1351–1363.

Upadhyay, Ashish Kumar, Abhishek Singh, Kaushalendra Kumar, and Ashish Singh. 2015. "Impact of Indoor Air Pollution from the Use of Solid Fuels on the Incidence of Life Threatening Respiratory Illnesses in Children in India." *BMC Public Health*, 15(1): 300–308.

WHO. 2000. *Addressing the Links between Indoor Air Pollution, Household Energy and Human Health*. World Health Organization.

WHO. 2014. *Indoor Air Quality Guidelines: Household Fuel Combustion*. World Health Organization.

WHO. 2016. *Burning Opportunity: Clean Household Energy for Health, Sustainable Development, and Wellbeing of Women and Children*. World Health Organization.

WHO. 2018a. *Burden of Disease from Household Air Pollution for 2016*. World Health Organization.

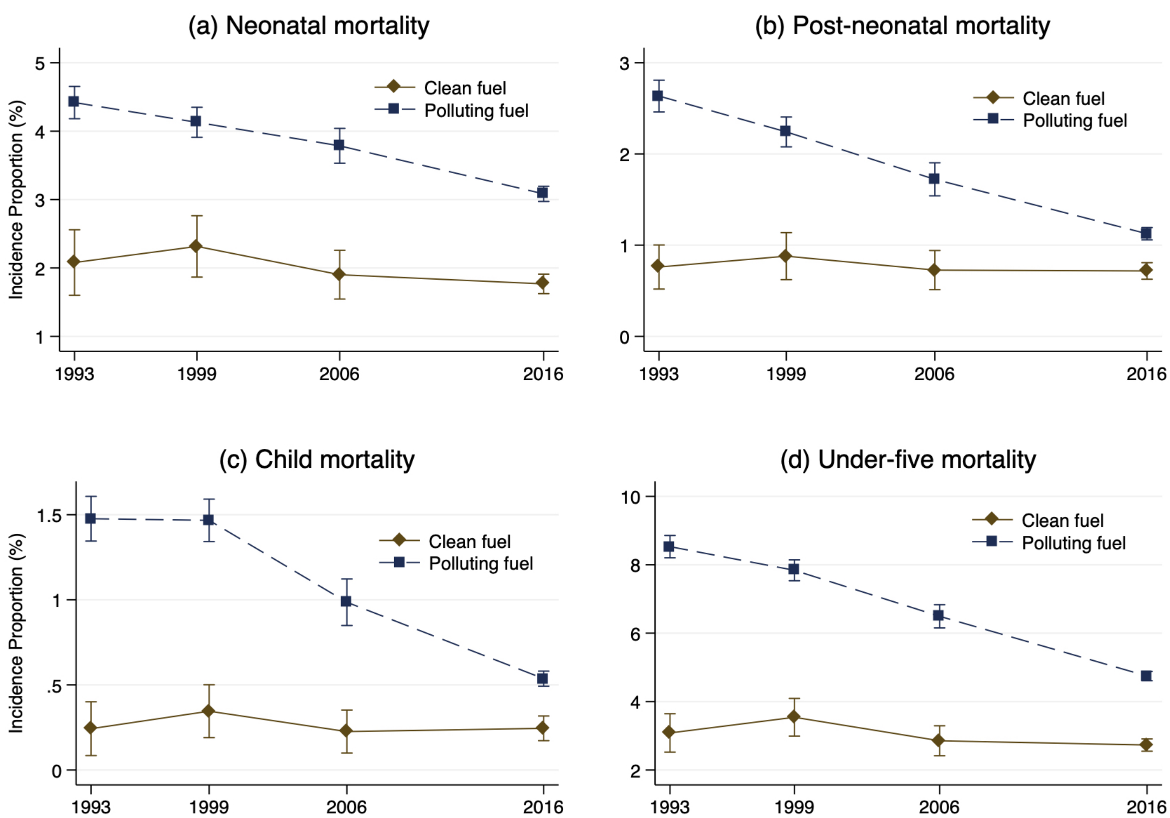
WHO. 2018b. *Household Air Pollution and Health, Fact Sheet No. 292*. World Health Organization.

WHO. 2018c. *World Health Statistics 2018: Monitoring Health for the SDGs*. World Health Organization.

Wichmann, J., and K.V.V. Voyi. 2006. "Influence of Cooking and Heating Fuel Use on 1–59 Month Old Mortality in South Africa." *Maternal and Child Health Journal*, 10(6): 553–561.

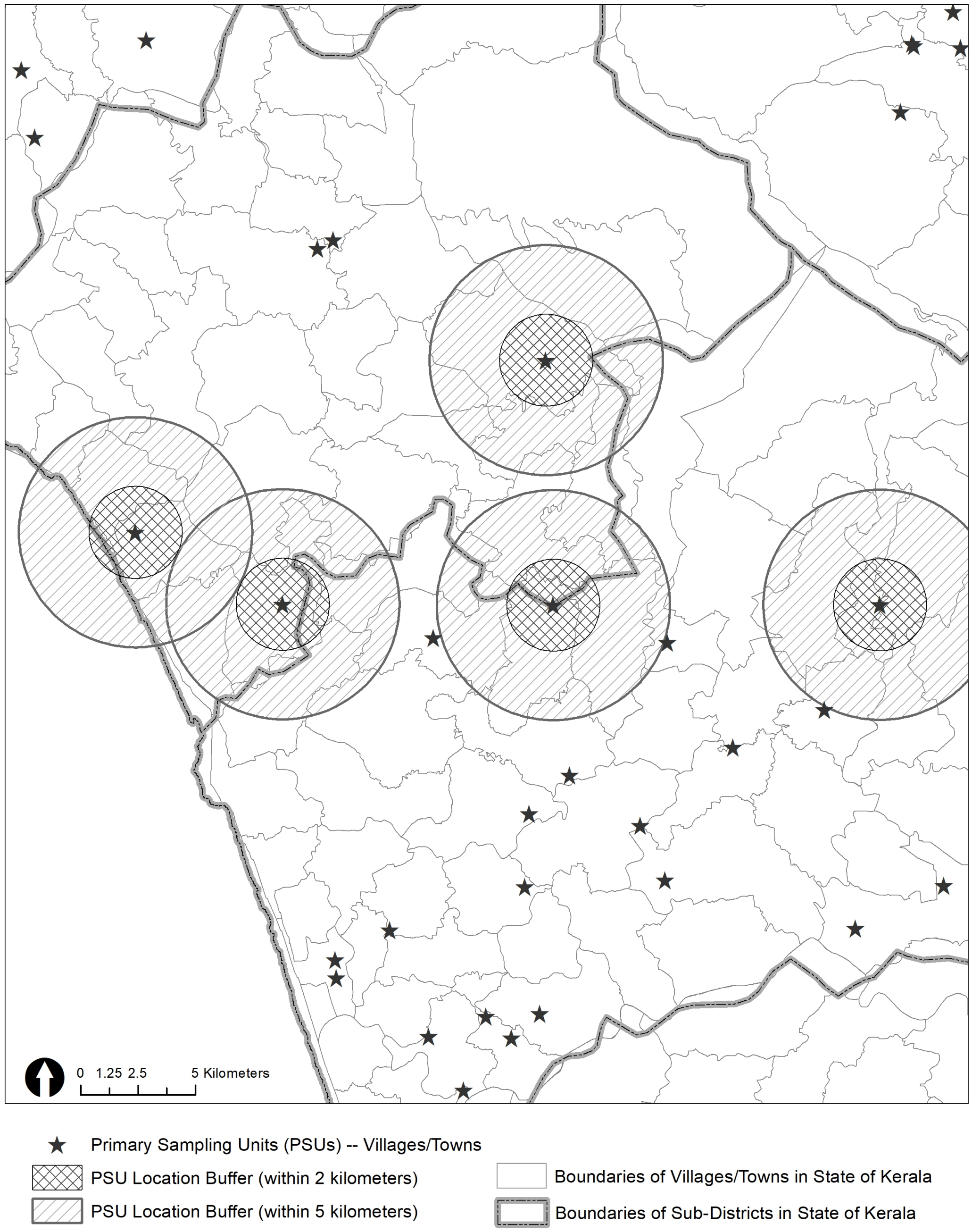
Zhang, Junfeng, and Kirk R. Smith. 2007. "Household Air Pollution from Coal and Biomass Fuels in China: Measurements, Health Impacts, and Interventions." *Environmental Health Perspectives*, 115(6): 848–855.

Figure 1: Mortality Trend in All Age-Groups of Children Under-Five by Cooking Fuel Type in India



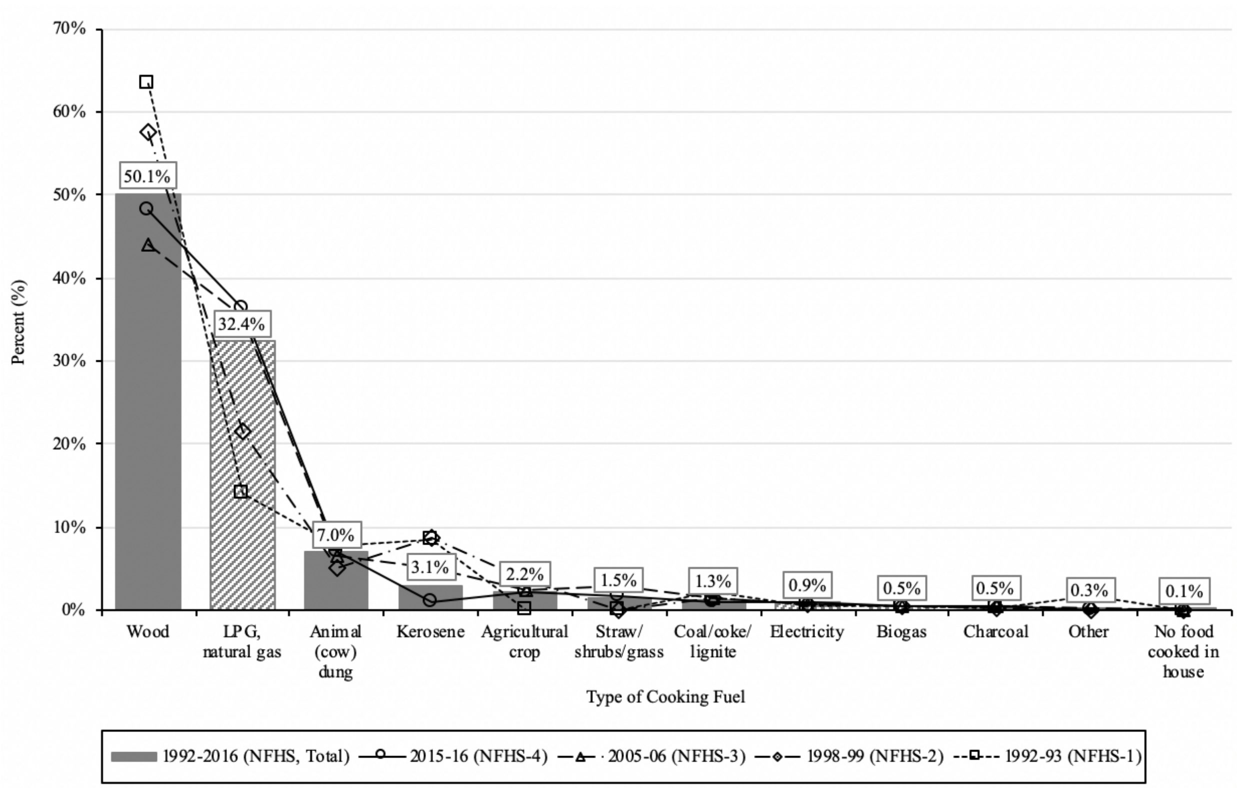
Notes: Based on NFHS datasets 1992–93, 1998–99, 2005–06, and 2015–16. In the medical literature, the measure of incidence proportion (or cumulative incidence) is defined as the proportion of individuals alive at the start of a period who die over that period (Greenland and Rothman, 2008; Centers for Disease Control and Prevention, 2006). To adjust for the cluster sampling survey design and apply the complex sample design parameters in estimating indicators, "svyset" and "svy" commands were used for calculating weighted estimates of mortality incidence proportion. The NFHS sample was selected through a two-stage sample design, and the commands deal with multiple stages of clustered sampling. Notice that the incidence proportions of neonatal, post-neonatal and child mortality add up to under-five mortality incidence because (i) these three preceding and successive age groups fully make up the first five years of life, and (ii) the measure of mortality incidence for all four different age groups have been calculated using a common denominator (or total number of live births).

Figure 2: Displacement of PSUs (Villages/City Blocks) in India's NFHS-4 (2015–16)



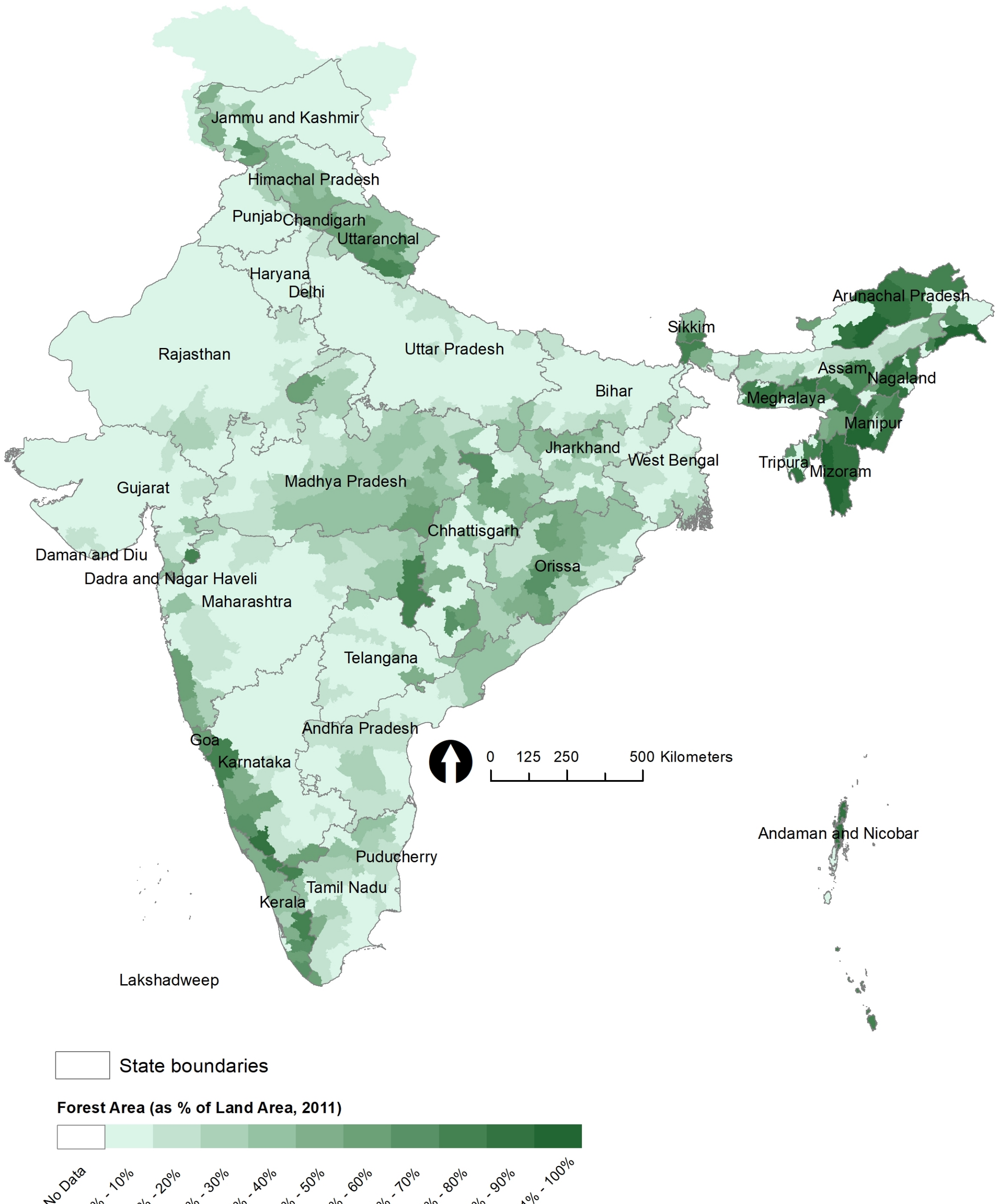
Notes: The figure shows how the PSU points are displaced in NFHS-4 (2015–16) survey based on few PSU points in Kerala district. In order to ensure that respondent confidentiality is maintained, the GPS (latitude/longitude positions) of respondent locations are randomly displaced according to the "random direction, random distance" method. The displacement is randomly carried out so that (i) urban clusters are displaced up to 2 kilometers, (ii) rural clusters are displaced up to 5 kilometers, with 1% of the rural clusters displaced up to 10 kilometers. According to the description of the DHS GPS data provided by the DHS Program, the displacement is restricted so that the points stay within the same country, state, and district areas as the undisplaced cluster. The buffer analysis on few PSU points in Kerala district as an example suggests that identification of villages/towns and sub-districts (or *tehsils*) is questionable because 2-5-kilometer buffers intersect with boundaries of villages/towns and sub-districts.

Figure 3: Share of Households in the NFHS relying on Different Types of Fuels for Cooking



Notes: The figure shows the share of households covered in four rounds of National and Family Health Survey (NFHS) using different types of fuels for cooking in India over the period 1992–2016. The line charts depict the share of households using each type of cooking fuel for each individual rounds of survey, while the bar chart illustrates the share for all four rounds of survey between 1992 and 2016 (the bars for clean fuels are filled with pattern, whereas the bars for polluting fuels are in solid fill). Wood is the leading fuel used for cooking in India, accounting for 50.1% of the sampling households in the NFHS over the period. The second dominant cooking fuel is a liquid petroleum gas (LPG) and/or natural gas with a share of 32.4%. The other clean fuels account for only 1.4% (electricity and biogas account for about 0.9% and 0.4%, respectively). Overall, based on our classification of cooking fuels, we can see that one-third of the Indian households have been consuming clean fuels for their cooking, while the majority or the remaining two-thirds of the households have been relying on polluting fuels for cooking over the past 25 years.

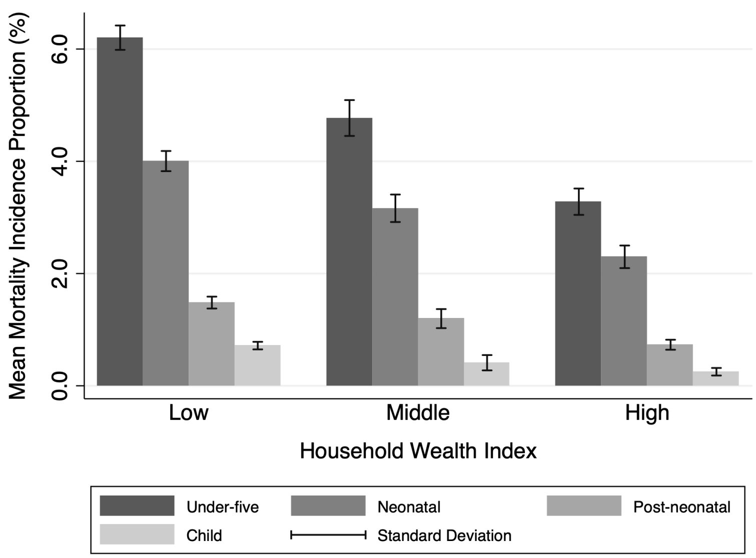
Figure 4: India's District-Wise and Satellite-based Forest Cover



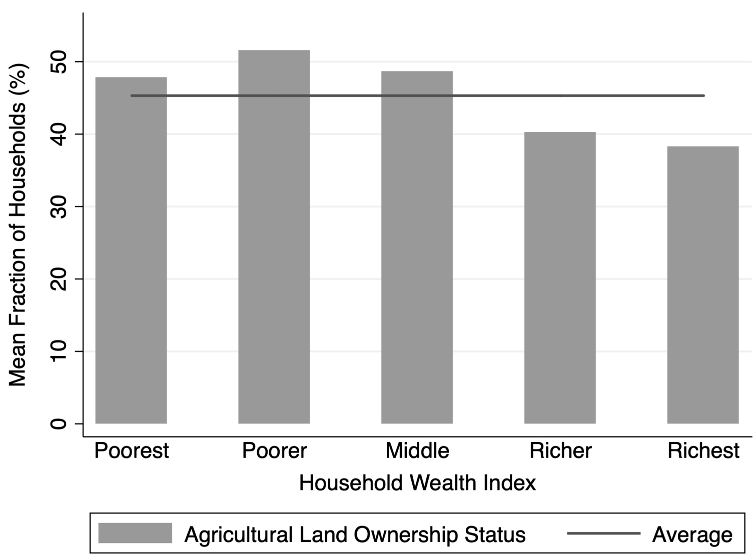
Notes: Based on a satellite-based data on forest cover from the Planning Commission of India. The figure depicts the 2011 district-wise forest cover (measured by percentage of geographical area covered by forests) in India. The forest cover includes all types of forests (different canopy density classes) including very dense (lands with tree canopy density of 70% and above), moderately dense (lands with tree canopy density between 40% and 70%), and open forests (lands with tree canopy density between 10% and 40%). The scrub or degraded forest lands with canopy density less than 10% is not considered for calculating forest cover.

Figure 5: Mortality in All Age-Groups of Children Under-Five and Ownership Status of Agricultural Land by Household Wealth in India

(a) Infant Mortality



(b) Ownership of Agricultural Land



Notes: Based on NFHS datasets 1992–93, 1998-99, 2005–06, and 2015–16. Panel (a) shows that under-five mortality incidence proportion is higher in households with lower wealth, suggesting that the probability of mortality decreases as a family becomes wealthier. The "low wealth" is the bottom 40% of households, "middle wealth" is the middle 40% of households, and "high wealth" is the top 20% of households. Panel (b) depicts the mean fraction of households that own land for agricultural purposes by dividing agricultural households into five groups (quintiles) based on household wealth.

Table 1: Summary Statistics

Variables	Mean	S.D.	Min	Max	*N*
Infant mortality (% total live births)					
Under-five	0.053	0.224	0.000	1.000	369,416
Child	0.007	0.085	0.000	1.000	369,416
Post-neonatal	0.015	0.120	0.000	1.000	369,416
Neonatal	0.031	0.173	0.000	1.000	369,416
Type of cooking fuel (% households)					
Clean	0.240	0.427	0.000	1.000	354,161
Polluting	0.760	0.427	0.000	1.000	354,161
Place of residence (% households)					
Urban	0.244	0.430	0.000	1.000	369,416
Rural	0.756	0.430	0.000	1.000	369,416
Household wealth (wealth index, % households)					
High	0.150	0.357	0.000	1.000	369,416
Middle	0.385	0.487	0.000	1.000	369,416
Low	0.465	0.499	0.000	1.000	369,416
Mother's age (years, % households)					
40-49	0.027	0.162	0.000	1.000	369,416
<20	0.041	0.199	0.000	1.000	369,416
20-29	0.679	0.467	0.000	1.000	369,416
30-39	0.253	0.435	0.000	1.000	369,416
Mother's education (% households)					
Secondary/Higher	0.458	0.498	0.000	1.000	369,219
Primary	0.151	0.358	0.000	1.000	369,219
No education	0.392	0.488	0.000	1.000	369,219
Gender of child (% households)					
Female	0.481	0.500	0.000	1.000	369,416
Male	0.519	0.500	0.000	1.000	369,416
Breastfeeding status (% households)					
Ever breastfed	0.654	0.476	0.000	1.000	369,416
Never breastfed	0.346	0.476	0.000	1.000	369,416
Place where food is cooked (% households)					
In same room as they live in	0.369	0.483	0.000	1.000	253,670
In separate kitchen inside the house	0.447	0.497	0.000	1.000	253,670
In a separate building	0.106	0.307	0.000	1.000	253,670
Outdoors	0.078	0.268	0.000	1.000	253,670
Type of house (% households)					
Pucca	0.376	0.484	0.000	1.000	358,410
Semi-pucca	0.437	0.496	0.000	1.000	358,410
Kachha	0.187	0.390	0.000	1.000	358,410
Number of household members	6.864	3.253	1.000	46.000	369,416

Notes: The table summarizes the household and individual characteristics of respondents from the three rounds of NFHS (1992–93, 1998–99, and 2015–16) used in the regression analysis. The unit of observation is the child. Neonatal = first 28 days of life (0–28 days), Post-neonatal = period between approximately the first month after birth and end of the first year of life (1-12 months), and Child = period between exact ages of one and five (12–59 months). Units are % household unless otherwise specified. The type of cooking fuel recorded in the survey as "no food cooked in house", "other", and "not a de jure resident" has been coded as missing observations.

Table 2: Summary Statistics of Infant Mortality and Fuel Choice (by State)

Panel A. Infant Mortality (fraction)

States	Under-Five		Child		Post-Neonatal		Neonatal		N
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	
Uttar Pradesh	0.074	0.262	0.010	0.099	0.021	0.142	0.044	0.204	56,090
Madhya Pradesh	0.066	0.249	0.012	0.108	0.017	0.129	0.038	0.190	32,007
Odisha	0.060	0.237	0.007	0.083	0.019	0.136	0.034	0.181	16,192
Rajasthan	0.059	0.235	0.009	0.096	0.017	0.129	0.033	0.177	25,435
Bihar	0.059	0.236	0.008	0.088	0.014	0.118	0.037	0.190	33,093
Assam	0.058	0.234	0.008	0.092	0.016	0.127	0.033	0.179	14,393
Chhattisgarh	0.058	0.235	0.006	0.08	0.012	0.110	0.040	0.196	10,695
Andhra Pradesh	0.055	0.228	0.006	0.076	0.016	0.125	0.033	0.179	5,515
Meghalaya	0.050	0.218	0.008	0.086	0.018	0.132	0.025	0.156	6,261
Gujarat	0.050	0.218	0.008	0.089	0.013	0.112	0.029	0.168	12,077
All States/UTs	0.053	0.224	0.007	0.085	0.015	0.120	0.031	0.173	369,416

Panel B. Type of Cooking Fuel (fraction)

States	Mean		S.D.	N
	Polluting	Clean		
Bihar	0.905	0.095	0.294	31,573
Meghalaya	0.896	0.104	0.305	6,247
Jharkhand	0.890	0.110	0.312	12,712
Odisha	0.881	0.119	0.323	15,487
West Bengal	0.865	0.135	0.342	9,033
Tripura	0.856	0.144	0.351	2,500
Assam	0.855	0.145	0.353	14,229
Chhattisgarh	0.838	0.162	0.368	10,108
Nagaland	0.829	0.171	0.376	5,646
Madhya Pradesh	0.812	0.188	0.391	30,658
All States/UTs	0.760	0.240	0.427	354,161

Notes: The table summarizes the infant mortality of four different age-groups (outcome variables, Panel A) and the type of cooking fuel (key explanatory variable, Panel B) by state recorded in three rounds of NFHS (1992–93, 1998–99, and 2015–16) used in the regression analysis. All 35 regions of India (29 states and six union territories–UTs) are considered, and we show 10 states/UTs with highest incidence of child mortality and highest share of households that use polluting fuel for cooking. Infant mortality and fuel choices significantly vary across regions throughout the country. In addition, six of these ten states/UTs (Odisha, Madhya Pradesh, Bihar, Assam, Chhattisgarh, and Meghalaya) are common in terms of highest fraction of polluting fuel use and under-five mortality incidence proportion.

Table 3: Summary Statistics of Infant Mortality and Fuel Choice (by Age and Gender of the Household Head)

	Infant mortality (fraction)									Type of cooking fuel (fraction)			
	Under-Five		Child		Post-Neonatal		Neonatal		N	Mean		S.D.	N
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.		Polluting	Clean		
Panel A. By Age of the Household Head													
Age 10-19	0.100	0.301	0.012	0.107	0.041	0.197	0.048	0.215	518	0.931	0.069	0.253	494
Age 20-29	0.058	0.235	0.007	0.082	0.016	0.125	0.036	0.186	62,807	0.812	0.188	0.391	62,067
Age 30-39	0.052	0.222	0.009	0.095	0.015	0.123	0.028	0.164	111,513	0.751	0.249	0.432	110,103
Age 40-49	0.061	0.239	0.010	0.099	0.017	0.127	0.034	0.182	53,663	0.791	0.209	0.407	50,169
Age 50-59	0.049	0.217	0.005	0.072	0.013	0.114	0.031	0.173	60,828	0.732	0.268	0.443	55,654
Age 60-69	0.046	0.209	0.005	0.069	0.012	0.109	0.029	0.168	56,211	0.720	0.280	0.449	52,830
Age 70-79	0.048	0.213	0.006	0.077	0.013	0.111	0.029	0.168	18,920	0.737	0.263	0.440	18,112
Age 80-89	0.052	0.221	0.006	0.076	0.016	0.126	0.030	0.170	4,254	0.761	0.239	0.427	4,071
Age $\geq$ 90	0.070	0.256	0.008	0.088	0.020	0.141	0.042	0.201	640	0.797	0.203	0.403	601
Total	0.053	0.224	0.007	0.085	0.015	0.120	0.031	0.173	369,354	0.760	0.240	0.427	354,101
Panel B. By Gender of the Household Head													
Male	0.053	0.225	0.008	0.086	0.015	0.121	0.031	0.174	330,884	0.762	0.238	0.426	317,867
Female	0.048	0.214	0.006	0.077	0.014	0.117	0.028	0.166	38,530	0.738	0.262	0.440	36,292
Total	0.053	0.224	0.007	0.085	0.015	0.120	0.031	0.173	369,414	0.760	0.240	0.427	354,159

Notes: The table summarizes the infant mortality and the type of cooking fuel for different age groups (Panel A) and gender (Panel B) of the household head recorded in three rounds of NFHS (1992–93, 1998–99, and 2015–16) used in the regression analysis. Note that household heads who are older than 50 years may have children with under five years of age. Intuitively, the number of children who live in households with heads older than 50 years declines as age of the household head increases.

Table 4: Probit: The Marginal Impact of Cooking Fuel Choice on Under-Five Mortality

	Dependent variable: Under-five mortality		
	(1)	(2)	(3)
Polluting fuel for cooking	0.023***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)
Place of residence: Rural		0.004***	0.004***
		(0.001)	(0.001)
Household wealth: Middle		0.012***	0.011***
		(0.002)	(0.002)
Household wealth: Low		0.015***	0.014***
		(0.002)	(0.002)
Number of household members		-0.004***	-0.004***
		(0.000)	(0.000)
Mother's age: $<$20		0.017***	0.017***
		(0.003)	(0.003)
Mother's age: 20-29		-0.009***	-0.008***
		(0.002)	(0.002)
Mother's age: 30-39		-0.013***	-0.013***
		(0.002)	(0.002)
Mother's education: Primary		0.008***	0.008***
		(0.001)	(0.001)
Mother's education: No education		0.010***	0.010***
		(0.001)	(0.001)
Gender of child: Male		0.004***	0.004***
		(0.001)	(0.001)
Never breastfed		0.049***	0.049***
		(0.001)	(0.001)
Food cooked: In separate kitchen inside		-0.002**	-0.003**
		(0.001)	(0.001)
Food cooked: In a separate building		-0.003	-0.003*
		(0.002)	(0.002)
Food cooked: Outdoors		0.000	0.000
		(0.002)	(0.002)
House type: Semi-pucca		0.004***	0.004***
		(0.001)	(0.001)
House type: Kachha		0.005**	0.005**
		(0.002)	(0.002)
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
State-by-Year FE	No	No	Yes
N	354,161	230,091	230,091
Probit log-likelihood	-71,407	-37,102	-37,094

Notes: Each column reports AMEs for a multivariate probit regression where the dependent variable is under-five mortality and the key explanatory variable is polluting fuel for cooking. The year fixed effects in Columns (2) and (3) include dummies for two years of interview (2015 and 2016). The state fixed effects include dummies for 36 states. The unit of observation is the child. Standard errors of the probit regressions are clustered at the PSU level, and standard errors of the AMEs in parentheses are computed by the delta method. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 5: Probit: The Marginal Impact of Cooking Fuel Choice on Child Mortality

	Dependent variable: Child mortality		
	(1)	(2)	(3)
Polluting fuel for cooking	0.007***	0.002***	0.002***
	(0.000)	(0.001)	(0.001)
Place of residence: Rural		0.001	0.001
		(0.000)	(0.000)
Household wealth: Middle		0.001	0.001
		(0.001)	(0.001)
Household wealth: Low		0.002*	0.002*
		(0.001)	(0.001)
Number of household members		-0.001***	-0.001***
		(0.000)	(0.000)
Mother's age: <20		-0.004***	-0.004***
		(0.001)	(0.001)
Mother's age: 20-29		-0.004***	-0.004***
		(0.001)	(0.001)
Mother's age: 30-39		-0.002***	-0.002***
		(0.001)	(0.001)
Mother's education: Primary		0.002***	0.002***
		(0.000)	(0.000)
Mother's education: No education		0.003***	0.003***
		(0.000)	(0.000)
Gender of child: Male		-0.001***	-0.001***
		(0.000)	(0.000)
Never breastfed		0.004***	0.004***
		(0.000)	(0.000)
Food cooked: In separate kitchen inside		-0.001*	-0.001*
		(0.000)	(0.000)
Food cooked: In a separate building		-0.001	-0.001
		(0.001)	(0.001)
Food cooked: Outdoors		-0.000	-0.000
		(0.001)	(0.001)
House type: Semi-pucca		0.001**	0.001**
		(0.000)	(0.000)
House type: Kachha		0.001*	0.001*
		(0.001)	(0.001)
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
State-by-Year FE	No	No	Yes
N	353,064	228,420	228,420
Probit log-likelihood	-14,927	-6,349	-6,349

Notes: Each column reports AMEs for a multivariate probit regression where the dependent variable is child mortality and the key explanatory variable is polluting fuel for cooking. The year fixed effects (FEs) in Columns (2) and (3) include dummies for two years of interview (2015 and 2016). The state fixed effects include dummies for 36 states. The number of observations is lower than that in Table 4 because there exist five states for which state FEs perfectly explain child mortality, and thus those five state FEs are dropped because probit models cannot be estimated when the outcome variable is perfectly predicted by the regressor. The unit of observation is the child. Standard errors of the probit regressions are clustered at the PSU level, and standard errors of the AMEs in parentheses are computed by the delta method. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 6: Probit: The Marginal Impact of Cooking Fuel Choice on Post-Neonatal Mortality

	Dependent variable: Post-neonatal mortality		
	(1)	(2)	(3)
Polluting fuel for cooking	0.007***	0.001	0.001
	(0.001)	(0.001)	(0.001)
Place of residence: Rural		0.002**	0.002**
		(0.001)	(0.001)
Household wealth: Middle		0.005***	0.005***
		(0.001)	(0.001)
Household wealth: Low		0.007***	0.007***
		(0.001)	(0.001)
Number of household members		-0.001***	-0.001***
		(0.000)	(0.000)
Mother's age: <20		0.003**	0.003**
		(0.002)	(0.002)
Mother's age: 20-29		-0.003***	-0.003***
		(0.001)	(0.001)
Mother's age: 30-39		-0.004***	-0.004***
		(0.001)	(0.001)
Mother's education: Primary		0.002***	0.002***
		(0.001)	(0.001)
Mother's education: No education		0.004***	0.004***
		(0.001)	(0.001)
Gender of child: Male		-0.000	-0.000
		(0.000)	(0.000)
Never breastfed		0.013***	0.013***
		(0.001)	(0.001)
Food cooked: In separate kitchen inside		-0.000	-0.000
		(0.001)	(0.001)
Food cooked: In a separate building		-0.000	-0.000
		(0.001)	(0.001)
Food cooked: Outdoors		0.000	0.000
		(0.001)	(0.001)
House type: Semi-pucca		0.000	0.000
		(0.001)	(0.001)
House type: Kachha		-0.000	-0.000
		(0.001)	(0.001)
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
State-by-Year FE	No	No	Yes
N	354,161	229,696	229,696
Probit log-likelihood	-26,484	-12,734	-12,734

Notes: Each column reports AMEs for a multivariate probit regression where the dependent variable is post-neonatal mortality and the key explanatory variable is polluting fuel for cooking. The year fixed effects (FEs) in Columns (2) and (3) include dummies for two years of interview (2015 and 2016). The state fixed effects include dummies for 36 states. The number of observations is slightly lower than that in Table 4 because there exists one state for which state FE perfectly explains post-neonatal mortality, and thus that state FE is dropped because probit models cannot be estimated when the outcome variable is perfectly predicted by the regressor. The unit of observation is the child. Standard errors of the probit regressions are clustered at the PSU level, and standard errors of the AMEs in parentheses are computed by the delta method. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 7: Probit: The Marginal Impact of Cooking Fuel Choice on Neonatal Mortality

	Dependent variable: Neonatal mortality		
	(1)	(2)	(3)
Polluting fuel for cooking	0.011***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
Place of residence: Rural		0.001	0.001
		(0.001)	(0.001)
Household wealth: Middle		0.006***	0.006***
		(0.001)	(0.001)
Household wealth: Low		0.007***	0.006***
		(0.002)	(0.002)
Number of household members		-0.002***	-0.002***
		(0.000)	(0.000)
Mother's age: <20		0.018***	0.018***
		(0.003)	(0.003)
Mother's age: 20-29		-0.001	-0.001
		(0.002)	(0.002)
Mother's age: 30-39		-0.007***	-0.007***
		(0.002)	(0.002)
Mother's education: Primary		0.004***	0.004***
		(0.001)	(0.001)
Mother's education: No education		0.003***	0.003***
		(0.001)	(0.001)
Gender of child: Male		0.005***	0.005***
		(0.001)	(0.001)
Never breastfed		0.032***	0.032***
		(0.001)	(0.001)
Food cooked: In separate kitchen inside		-0.002*	-0.002*
		(0.001)	(0.001)
Food cooked: In a separate building		-0.002	-0.002
		(0.001)	(0.001)
Food cooked: Outdoors		0.000	0.000
		(0.001)	(0.001)
House type: Semi-pucca		0.003***	0.003***
		(0.001)	(0.001)
House type: Kachha		0.004***	0.004***
		(0.002)	(0.002)
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
State-by-Year FE	No	No	Yes
N	354,161	230,091	230,091
Probit log-likelihood	-47,742	-26,355	-26,348

Notes: Each column reports AMEs for a multivariate probit regression where the dependent variable is neonatal mortality and the key explanatory variable is polluting fuel for cooking. The year fixed effects in Columns (2) and (3) include dummies for two years of interview (2015 and 2016). The state fixed effects include dummies for 36 states. The unit of observation is the child. Standard errors of the probit regressions are clustered at the PSU level, and standard errors of the AMEs in parentheses are computed by the delta method. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 8: The Effect of Polluting Fuel for Cooking on Infant Mortality
(Comparison of Results from Simple Logit Regressions)

	(1)	(2)	(3)	(4)	(5)
	NFHS-1-3 (1992–2006)		NFHS-4	NFHS-1-4	This
	Naz et al. (2016)	Replication	(2015–16)	(1992–2016)	paper
	Dependent variable: Under-five mortality				
Odds Ratio	1.30***	1.27***	1.32***	1.26***	1.26***
		(0.065)	(0.114)	(0.054)	(0.047)
Marginal Effect		0.014***	0.011***	0.013***	0.009***
		(0.003)	(0.003)	(0.003)	(0.001)
N	138,063	150,845	34,423	185,268	230,091
	Dependent variable: Child mortality				
Odds Ratio	1.42**	1.45**	0.99	1.24	1.46***
		(0.231)	(0.256)	(0.162)	(0.177)
Marginal Effect		0.004**	0.000	0.002	0.002***
		(0.002)	(0.001)	(0.001)	(0.001)
N	138,063	150,845	34,423	185,268	228,420
	Dependent variable: Post-neonatal mortality				
Odds Ratio	1.42***	1.42***	1.14	1.30***	1.08
		(0.136)	(0.187)	(0.105)	(0.073)
Marginal Effect		0.007***	0.001	0.004***	0.001
		(0.002)	(0.002)	(0.001)	(0.001)
N	138,063	150,845	34,423	185,268	229,696
	Dependent variable: Neonatal mortality				
Odds Ratio	1.23***	1.18**	1.46***	1.25***	1.30***
		(0.076)	(0.158)	(0.068)	(0.061)
Marginal Effect		0.006**	0.010***	0.007***	0.007***
		(0.002)	(0.003)	(0.002)	(0.001)
N	138,063	150,845	34,423	185,268	230,091

Notes: Column (1) shows the odds ratio from logit regression in Naz et al. (2016), while Columns (2), (3) and (4) show the odds ratio from logit regression with specification exactly the same as in Naz et al. (2016). The differences in odds ratio presented in Columns (1) and (2) are due to difference in number of observations because we control for exactly the same variables as in Naz et al. (2016) (including type of cooking fuel, place of residence, wealth index, mother's age, mother's education, mother's working status, sex of child, breastfeeding status, separate kitchen, type of house, and year of survey). We have very few observations in Column (3) because only a (state module) sub-sample of women were asked about their employment status, resulting in a lot of missing observations for mother's working status variable in the NFHS-4 (2015–16). Column (5) presents odds ratio and the associated average marginal effects from logit regressions with our primary specification (or specification in Column (3) of Tables 4-7). One of our controls, a variable indicating whether household cooks inside the house, in a separate building, or outdoors, is only available in NFHS-4, thus, we use only last round of the survey in Column (5). The numbers of observations for child and post-neonatal mortality regressions are lower than that for under-five and neonatal mortality regressions in Column (5) because there exist respectively five and one state(s) for which state FEs perfectly explain child and post-neonatal mortality, and thus those state FEs are dropped. It is because logit models cannot be estimated when the outcome variable is perfectly predicted by the regressor. The unit of observation is the child. Standard errors of the logit regressions in parentheses are clustered at the primary sampling unit (PSU) level, and standard errors of the corresponding AMEs in parentheses are computed by the delta method. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 9: Cooking Fuel Choice and Infant Mortality from IV (2SLS) Regressions
(IV = Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting Fuel Use	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.050***	0.003	0.010	0.037***
		(0.017)	(0.006)	(0.009)	(0.014)
Owns agricultural land	0.057***				
	(0.002)				
N	230,091	230,091	230,091	230,091	230,091
R^2	0.54	0.02	0.00	0.01	0.01
F-stat on IV	799.71				

Notes: All specifications contain an unreported vector of demographic controls and constant term. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The state, year, and state-by-year fixed effects are also included in every specification. The OLS model does not drop the state FE(s) that perfectly explain the child and post-neonatal mortality incidences, and thus the number of observations is the same across four IV regressions. The unit of observation is the child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

Table 10: Cooking Fuel Choice and Infant Mortality from OLS and IV (2SLS) Regressions (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
		Panel A. IV			
	1^{st} stage	2^{nd} stage			
Polluting fuel for cooking		0.049**	0.004	0.012	0.034**
		(0.020)	(0.007)	(0.010)	(0.015)
Speed of change in forest cover	0.030***				
	(0.011)				
Owns agricultural land	0.054***				
	(0.003)				
N	196,344	196,344	196,344	196,344	196,344
R^2	0.54	0.02	0.00	0.01	0.01
F-stat on IVs	169.10				
Hansen's J-statistic		0.92	1.92	0.74	1.80
χ^2 p-value		0.34	0.17	0.39	0.18
		Panel B. OLS			
Polluting fuel for cooking		0.008***	0.001***	0.001	0.006***
		(0.001)	(0.000)	(0.001)	(0.001)
N		230,091	230,091	230,091	230,091
R^2		0.03	0.00	0.01	0.02

Notes: The first column in Panel A reports result from the first-stage regression of our IV (2SLS) regression using NFHS-4 data. The dependent variable is a binary variable of whether fuel choice. The F-test on IVs—district-wise speed of change in forest cover calculated as a relative change in the percentage of forested area in the total geographical area of the region over the period 2007, 2011, and 2013 using satellite-based data and an indicator variable for household's agricultural land ownership—verifies that the instruments generate a plausible variation in polluting fuel for cooking. Columns (2), (3), (4) and (5) in Panel A report results from the estimation of Equation (1) using IV regression with different dependent variables and similar specification where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The Hansen's J-statistic suggests that the excluded IVs are exogenous. Panel B reports coefficient estimates of the association between use of polluting fuel for cooking on the infant mortality of four different age-groups. Coefficients are from the estimation of Equation (1) using OLS regression based on the same dataset used in Tables 4-7. Notice that we drop a few unmatched observations, which is less than 1 percent of our total observations, when we merge the survey dataset with the satellite and Census datasets, and it does not affect our primary results from the estimation of simple models. All specifications of OLS and IV regressions contain an unreported vector of demographic controls and constant term. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The state, year, and state-by-year fixed effects are also included in every specification. The unit of observation is the child. Heteroskedasticity-consistent standard errors clustered by districts are in parentheses. The statistical significances of the key regressors are the same for all simple OLS and IV (2SLS) regressions when the standard errors are clustered by PSUs. Significance: $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

Appendix

Table A.1: Cooking Fuel Choice and Infant Mortality from Pooled and IV Probit Regressions (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership)

	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
		Panel A. IV Probit			
	1^{st} stage	2^{nd} stage			
Polluting fuel for cooking		0.556***	0.248	0.439	0.569**
		(0.207)	(0.477)	(0.340)	(0.237)
Speed of change in forest cover	0.030***				
	(0.004)				
Owns agricultural land	0.054***				
	(0.002)				
N	196,344	196,344	195,212	195,949	196,344
R^2	0.54				
F-stat on IVs	339.59				
Model Wald χ^2		3,913.63	583.90	1,319.46	2,503.41
Model degrees of freedom		49.00	46.00	48.00	49.00
Model Wald p-value		0.00	0.00	0.00	0.00
Exogeneity test Wald p-value		0.03	0.78	0.25	0.05
Wald χ^2 test of exogeneity		4.76	0.08	1.34	3.78
		Panel B. Pooled Probit			
Polluting fuel for cooking		0.097***	0.127***	0.030	0.104***
		(0.017)	(0.041)	(0.026)	(0.020)
N		230,091	228,420	229,696	230,091
Probit log-likelihood		-37,094	-6,349	-12,734	-26,348

Notes: The first column in Panel A reports result from the first-stage OLS regression of IV probit using NFHS-4 where the dependent variable is a binary variable for polluting fuel. The F-test on IVs—district-wise speed of change in forest cover calculated as a relative change in the percentage of forested area in the total geographical area of the region over the period 2007, 2011, and 2013 using satellite-based data and an indicator variable for household's agricultural land ownership—confirms that the instruments create a significant variation in polluting fuel for cooking. Columns (2), (3), (4) and (5) in Panel A report coefficient estimates from the estimation of Equation (1) using IV probit regression with different dependent variables and a similar specification. Panel B reports coefficient estimates of the association between use of polluting fuel for cooking on the infant mortality of four different age-groups. Coefficients are from the estimation of Equation (1) using pooled probit regressions, which are exactly the same as those in Tables 4-7. All specifications of pooled and IV probit regressions contain an unreported vector of demographic controls and constant term. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The state, year, and state-by-year fixed effects (FEs) are also included in every specification. Some state FEs are excluded because they perfectly predict the outcome variable in child and post-neonatal mortality regressions. The unit of observation is the child. Heteroskedasticity-consistent standard errors clustered by PSUs are in parentheses. The standard errors of the key regressors in the first-stage regression and joint F-statistic on the excluded IVs are different from those in Column (1) of Table 10 due to difference in cluster level. However, the statistical significances of the key regressors are the same for all pooled and IV probit regressions when the standard errors are clustered by districts. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.2: Cooking Fuel Choice and Infant Mortality from IV (2SLS) Regressions
(IVs = Speed of Change in Forest Cover and Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting Fuel Use	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.049***	0.004	0.012	0.034**
		(0.019)	(0.006)	(0.010)	(0.015)
Cookstoves Program States (NBCI)	0.056	0.055***	0.006***	0.011**	0.038***
	(0.055)	(0.007)	(0.001)	(0.005)	(0.005)
Speed of change in forest cover	0.030***				
	(0.004)				
Owns agricultural land	0.054***				
	(0.002)				
N	196,344	196,344	196,344	196,344	196,344
R^2	0.54	0.02	0.00	0.01	0.01
F-stat on IVs	339.59				
Hansen's J-statistic		1.21	1.76	0.92	1.67
χ^2 p-value		0.27	0.18	0.34	0.20

Notes: The first column reports result from the first-stage regression of 2SLS regression using NFHS-4 where the dependent variable is a binary variable for polluting fuel. The F-test on IVs—district-wise speed of change in forest cover calculated as a relative change in the percentage of forested area in the total geographical area of the region over the period 2007, 2011, and 2013 using satellite-based data and an indicator variable for household's agricultural land ownership—verifies that the instruments generate a plausible variation in polluting fuel for cooking. Columns (2), (3), (4) and (5) report results from the second-stage regressions of 2SLS regression with different dependent variables and similar specification where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. All specifications contain an unreported vector of demographic controls and constant term. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. In addition to these demographic controls, we control for a dummy variable indicating states where the National Biomass Cookstove Initiative (NBCI) has been implemented by the government of India. The state, year, and state-by-year fixed effects are also included in every specification. The Hansen's J-statistic suggests that the excluded IVs are exogenous. The unit of observation is the child. Parentheses contain standard errors clustered by PSUs. Significance: $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

Table A.3: Levels of Dirtiness of Cooking Fuels and Infant Mortality from IV (2SLS) Regressions (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting Fuel Use	Under-Five	Child	Post-Neonatal	Neonatal
Dirtiness level of cooking fuels		0.009***	0.000	0.002	0.006**
		(0.003)	(0.001)	(0.002)	(0.002)
Speed of change in forest cover	0.066***				
	(0.025)				
Owns agricultural land	0.344***				
	(0.014)				
N	196,344	196,344	196,344	196,344	196,344
R^2	0.52	0.02	0.00	0.01	0.01
F-stat on IVs	313.55				
Hansen's J-statistic		0.24	2.01	0.44	0.61
χ^2 p-value		0.62	0.16	0.51	0.44

Notes: All specifications contain an unreported vector of demographic controls and constant term. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The state, year, and state-by-year fixed effects are also included in every specification. The Hansen's J-statistic suggests that the excluded IVs are exogenous. The unit of observation is the child. Parentheses contain standard errors clustered by PSUs. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.4: Cooking Fuel Choice and Infant Mortality from IV (2SLS) Regressions
(IVs = Satellite-based Forest Cover and Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Polluting fuel for cooking		0.046***	0.004	0.010	0.032**
		(0.018)	(0.006)	(0.009)	(0.014)
Forest cover (*satellite-based, 2011*)	0.050***				
	(0.008)				
Owns agricultural land	0.057***				
	(0.002)				
N	206,548	206,548	206,548	206,548	206,548
R^2	0.54	0.02	0.00	0.01	0.01
F-stat on IVs	376.95				
Hansen's J-statistic		1.24	1.90	1.01	1.70
χ^2 p-value		0.27	0.17	0.31	0.19

Notes: The first column reports result from the first-stage regression of our 2SLS regression using NFHS-4 data. The dependent variable is a binary variable of whether fuel choice. The F-test on IVs—2011 district-wise forest cover calculated as a percent of total geographical area of the region using satellite-based data and an indicator variable for household's agricultural land ownership—verifies that the instruments generate a plausible variation in polluting fuel for cooking. Columns (2), (3), (4) and (5) report results from the second-stage regressions of 2SLS regression with different dependent variables and similar specification where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. All specifications contain an unreported vector of demographic controls and constant term. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The state, year, and state-by-year fixed effects are also included in every specification. The Hansen's J-statistic suggests that the excluded IVs are exogenous. The unit of observation is the child. Parentheses contain standard errors clustered by PSUs. Significance: $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

Table A.5: Cooking Fuel Choice and Infant Mortality from IV (2SLS) Regressions (IVs = Census-based Forest Cover and Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting Fuel Use	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.036**	0.002	0.000	0.033**
		(0.018)	(0.006)	(0.009)	(0.014)
Forest cover (*census-based, 2011*)	0.037***				
	(0.009)				
Owns agricultural land	0.057***				
	(0.002)				
N	212,493	212,493	212,493	212,493	212,493
R^2	0.54	0.03	0.00	0.01	0.02
F-stat on IVs	385.00				
Hansen's J-statistic		2.34	0.59	3.72	1.16
χ^2 p-value		0.13	0.44	0.05	0.28

Notes: The first column reports result from the first-stage regression of 2SLS regression using NFHS-4 where the dependent variable is a binary variable for polluting fuel. The F-test on IVs—district-wise forest cover calculated as a percent of total geographical area of the region using the 2011 Indian Census and an indicator variable for household's agricultural land ownership—verifies that the instruments generate a plausible variation in polluting fuel for cooking. Columns (2), (3), (4) and (5) report results from the second-stage regressions of 2SLS regression with different dependent variables and similar specification where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. All specifications contain an unreported vector of demographic controls and constant term. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The state, year, and state-by-year fixed effects are also included in every specification. The Hansen's J-statistic suggests that the excluded IVs are exogenous. The unit of observation is the child. Parentheses contain standard errors clustered by PSUs. Significance: $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

OTHER A.E.M. WORKING PAPERS

WP No	Title	Fee (if applicable)	Author(s)
2020-04	Cooking Fuel Choice, Indoor Air Quality and Child Mortality in India		Basu, K. A., Byambasuren, T., Chau, N., and Khanna N.
2020-03	DIFFERENCES IN MILK PAYMENT STRUCTURE BY COOPERATIVE AND INDEPENDENT HANDLERS: An Examination from New York State		Munch, D. M., Schmit, T.M., and Severson, R.M.
2020-02	COVID-19 Impact on Fruit and Vegetable Markets		Richards, T., and Rickard, B.
2020-01	Date Labels, Food Waste, and Implications for Dietary Quality		Rickard, B., Ho, S.T., Livat, F., and Okrent, A.
2019-12	A Review of Economic Studies on Pathogen-Tested Plant Materials and Clean Plant Programs for Specialty Crops		Yeh, A. D., Park, K., Gomez, M., Fuchs, M.
2019-11	Short-Term and Long-Term Effects of Trade Liberalization		Lin, G. C.
2019-10	Using the Alternative Minimum Tax to Estimate the Elasticity of Taxable Income for Higher-Income Taxpapers		Abbas, A.
2019-09	In Praise of Snapshots		Kanbur, R.
2019-08	The Index Ecosystem and the Commitment to Development Index		Kanbur, R.
2019-07	Promoting Education Under Distortionary Taxation: Equality of Opportunity versus Welfarism		Haaparanta, P., Kanbur, R., Paukkeri, T., Pirttilä, J. & Tuomala, M.
2019-06	Management Succession Lessons Learned from Large Farm Businesses in Former East Germany		Staehr A. E.
2019-05	A Narrative on Two Weaknesses of the TRI for Research Purposes		Khanna N.
2019-04	Village in the City: Residential Segregation in Urbanizing India		Bharathi N., Malghan D., Rahman A.
2019-03	Inequality in a Global Perspective		Kanbur R.
2019-02	Impacts of Minimum Wage Increases in the U. S. Retail Sector: Full-time versus Part Time Employment		Yonezawa K., Gomez M., McLaughlin M.,
2019-01	Minimum Wages and Labor Supply in an Emerging Market: the Case of Mauritius		Asmal Z., Bhorat H., Kanbur R., Ranzani M., Paci P.

Paper copies are being replaced by electronic Portable Document Files (PDFs). To request PDFs of AEM publications, write to (be sure to include your e-mail address): Publications, Department of Applied Economics and Management, Warren Hall, Cornell University, Ithaca, NY 14853-7801. If a fee is indicated, please include a check or money order made payable to Cornell University for the amount of your purchase. Visit our Web site (*http://dyson.cornell.edu/research/wp.php)* for a more complete list of recent bulletins.