Applied Economics and Policy Working Paper Series

AEP WP NO. 2021-10

CHILD MORALITY AND INDOOR AIR POLLUTION

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

Abstract

How serious is indoor air pollution (IAP) a mortality threat to young children? This paper estimates the causal effect of cooking fuel choice – a predominant cause of IAP – on infant mortality in India (1992-2016), where the most health-endangering biomass fuels are also most commonplace. Leveraging the speed of change in forest cover and land ownership for identification, we find polluting fuel choice to impose highly heterogeneous local infant mortality effects by age group (from insignificant to 4:7 percent increase) – implying the loss of two lives every 1,000 live births. These conclusions are robust to alternative estimation strategies and additional controls.

Charles H. Dyson School of Applied Economics and Management

School of Hotel Administration

Samuel Curtis Johnson Graduate School of Management

Cornell
SC Johnson College of Business

Child Mortality and Indoor Air Pollution*

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

This version: July 15, 2021

Abstract

How serious is indoor air pollution (IAP) a mortality threat to young children? This paper estimates the causal effect of cooking fuel choice – a predominant cause of IAP – on infant mortality in India (1992-2016), where the most health-endangering biomass fuels are also most commonplace. Leveraging the speed of change in forest cover and land ownership for identification, we find polluting fuel choice to impose highly heterogeneous local infant mortality effects by age group (from insignificant to 4.7 percent increase) – implying the loss of two lives every 1,000 live births. These conclusions are robust to alternative estimation strategies and additional controls.

JEL Classification: I18, N35, Q53.

Keywords: child mortality, indoor air pollution, cooking fuel, 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, 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

For young children and their caretakers in developing countries, indoor air pollution (IAP) is a silent health and mortality threat. How many lives can potentially be saved if indoor air quality is improved? IAP is widely identified as globally the most health-endangering environmental factor (World Health Organization, 2016). More than 40% of the world population – almost 3 billion individuals – rely on open fire or simple stoves fueled by dirty fuels including coal, biomass fuel, and kerosene. Of these, a vast majority burn organic biomass fuels (e.g., wood, agricultural waste, and animal dung) for cooking and as domestic sources of energy (World Health Organization, 2018a). Around 95% of these individuals live in poverty in the low and middle-income countries of Asia, Western Pacific, and Africa (Duflo et al., 2008b).[1]

IAP is primarily due to incomplete burning of solid fuels for heating, lighting, and cooking. The combination of traditional cooking stoves and polluting fuels generates high levels of hazardous indoor air pollutants (e.g., Fullerton et al., 2008; Duflo et al., 2008b; Rehfuess et al., 2011; Barron and Torero, 2017). Diseases attributed to poor air quality inside the house (e.g., heart disease, respiratory diseases) are the leading causes of death worldwide, to the tune of 4 million people per year, (World Health Organization, 2018a)[2] and the biggest cause of disability-adjusted life years (DALYs) lost in Southeast Asia and Sub-Saharan Africa.[3] Globally, IAP ranks third place among the lists of causes of DALYs lost (Apte and Salvi, 2016).

Women and young children (especially, children younger than five years old) are particularly vulnerable to the negative health risks associated with IAP, as women often disproportionately and simultaneously shoulder the responsibility of cooking and child care (Edwards and Langpap, 2012). For example, in India, approximately 56% of under-five children stay with their mothers at all times including during cooking (Rehfuess et al., 2011). Acute lower respiratory infections (ALRI), including pneumonia, is the second dominant cause of under-five mortality worldwide after premature birth, and one-third of ALRI-related deaths are said to be due to poor air quality at home (World Health

[1] This includes 80% of the population in China, 82% in India, 87% in Ghana, 95% in Afghanistan, and 95% in Chad who rely primarily on polluting cooking fuels.

[2] According to World Health Organization (2016), IAP from the consumption of polluting fuels led to 3.8 million premature mortality in 2016 alone, or 6.7% of global deaths, which is more serious than total death tolls of tuberculosis, malaria and HIV/AIDS. Of these, 403,000 were under-five children. 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).

[3] The DALY is the most commonly used measure of national burden of disease and combines years lost because of disability with those because of death.

Organization, 2018b).[4] We found nearly 200 epidemiological publications that study the health consequences of solid fuels, e.g., biomass fuels and coal (Zhang and Smith, 2007). These provided some of the first systematic evidence of the correlation of adverse health outcomes and IAP including chronic obstructive pulmonary disease (COPD), ALRI, asthma, lung cancer, and immune system impairment (Zhang and Smith, 2007). In parallel, Duflo et al. (2008a) presents one of the earliest studies on the health impact of IAP, and finds a high degree of correlation between using a traditional stove and having symptoms of respiratory illness using a linear probability model with a variety of controls.

With respect to the mortality implications of IAP, Naz et al. (2016) finds an odds ratio of 1.30 between indoor polluting fuel use and under-five mortality in India using household survey data. This is before any measures are undertaken to address potential endogeneity and omitted variable biases (e.g., household size, dwelling size, local governance and cultural factors, distance to towns, and ambient air and soil quality).[5] The reliability of these and other early non-causal estimates of the consequences of IAP have been questioned due to inadequate controls for health outcomes and lack of convincing identification strategies (Duflo et al., 2008b). Given that the youngest and most vulnerable members of the household are more homebound, a shift in cooking fuel choice practice to improve indoor air quality carries with it the potential to impact the health and long term well-being of the lion's share of households in the developing world even without resorting to large scale intervention in infrastructure. The validity of the IAP-infant link, and its causal nature, in particular, beg investigation.

[4]There are a number of studies on IAP and health outcomes. Silwal and McKay (2015) find that the use of firewood instead of kerosene, LPG and electricity for cooking damages lung capacity by 9.4 percent in Indonesia using proximity to the nearest market to instrument for fuel choice. 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 as instruments for the two regressors of interest. Both of these studies find weak justification in the tests to validate the exclusion restriction and exogeneity of instruments, however. Pitt et al. (2006) examine the effect of time spent cooking on incidence of any respiratory symptom for all adults and adult women using gender-specific hierarchies as instruments for exposure to IAP. They find that a four hour per day increase in the time spent cooking increases the likelihood of having a respiratory symptom by 10.8 percentage points. Liu et al. (2020) investigate the causal effect of using non-solid fuels instead of solid fuels on elder's health by estimating the capability to deal with activities of daily living (ADL) and instrumental ADL on household fuel choice in rural China using the share of village citizens who rely on clean fuels for cooking as an instrument. These studies do not examine the role of IAP on infant mortality, however.

[5]For instance, the demographic and political characteristics of the state governors, such as their age, education, gender, 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. Region-specific socio-cultural trends may also prevent households from switching to clean energy, for example when rural households consider the use of animal dung as clean and natural in addition to lower cost. Distance to towns can proxy ease of access to clean fuels, including electricity for example. Ambient air quality may reflect local abundance of forests and coal deposits, and while soil quality can proxy for agricultural land use and agricultural crop waste.

There is now a small and growing literature on the causal health impact of exposure to IAP. The findings so far are nuanced. The first randomized control trial (RCT) experiments aimed at demonstrating health impacts of IAP were carried out in San Marcos city, Guatemala (Diaz et al., 2007; Smith-Sivertsen et al., 2009). Using logistic random intercept models, the study finds 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. Contrary to these findings, the longer term impacts of improved cook stoves have appeared less promising. To this end, Hanna et al. (2016) demonstrates that improved cook stoves, in fact, did not reduce smoke exposure following the second year of installation, or improve the health of recipients and greenhouse gas emissions at all in an RCT in rural Orissa, India, due to improper use or lack of maintenance.[6]

In addition to dedicated RCTs, only a few related national level studies in developing economies exist, but these focus almost exclusively on outdoor air quality, leveraging, for example, intertemporal and spatial heterogeneity in the incidences of wildfires, meteorological shocks, exogenous shifts in national energy infrastructure, industrial structure/technology and cross-border pollution for identification (Jayachandran, 2009; Arceo et al., 2016; Cesur et al., 2017; Beach and Hanlon, 2018; Benshaul-Tolonen, 2019; Jia and Ku, 2019). An exception is Imelda (2018), which uses a quasi-experimental difference-in-differences strategy to estimate that a 1.1 percent reduction in infant mortality rate or four saved infant lives every 10,000 live births[7] resulted from an Indonesian government program of subsidizing households to switch from kerosene to liquid petroleum gas (LPG) as cooking fuel.[8] No studies to date have presented causal estimates

[6]Other studies that focus 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 include Bruce et al. (2004), Duflo et al. (2008a), Smith-Sivertsen et al. (2009), Hanna et al. (2016), and Barron and Torero (2017).

[7]We consider infant mortality, similarly defined as in Imelda (2018), as an alternative outcome variable to compare implication of our analysis with that of Imelda (2018) and find that a family using polluting fuels for cooking has a 4.8 percent higher probability of experiencing infant mortality within one year of birth, implying the loss of 25 lives every 10,000 live births. Hence, the reason why we have a higher estimate on the death toll of IAP is that we include, for example, biomass fuels in our list of polluting fuels, which are the most dangerous for the life or health of a child.

[8]In a follow-up paper, Imelda (2020) finds a 16-34 percent and an 8-25 percent reduction respectively in infant mortality, similarly defined as in Imelda (2018), and child's birth weight response to the same government program with a highlight on its effects on perinatal mortality—deaths within the first week after birth—using difference-in-differences (DID) specification with district-specific time trends. The author also shows that the intensity of the program leads to a 1.2 percentage point reduction in both infant and perinatal mortality with no impact on postneonatal mortality—deaths during the period 1-12 months of life—by replacing a binary treatment variable with a continuous variable of program intensity in the DID equation.

of the impact of polluting fuel choice that account for the role of biomass fuel despite its overwhelming popularity in developing countries.

This paper contributes to the nascent literature on the IAP-child mortality link by estimating the causal impact of indoor air pollution through the use of polluting fuels for cooking on under-five mortality. Our analysis is based on a large-scale household survey in India that contain records of household-level health and demographic information, as well as 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, also referred to as the Demographic and Health Survey–DHS) — NFHS-1 (1992–93), NFHS-2 (1998–99), and NFHS-4 (2015–16) with detailed observations on 369,416 singleton live-born children, of whom 19,474 died in the 5-years prior to the respective survey years. An important reason why we chose India as a starting point here is that the NFHS survey includes detail questions regarding respondents' use of 10 possible types of cooking fuels, including biomass fuels (firewood, straw, shrubs, grass, agricultural waste, and dung), kerosene, coal/lignite, charcoal, as well as biogas, LPG and natural gas, and electricity. The vast majority of Indian families rely on biomass fuels for domestic uses such as cooking. In this regard, we see India as a good representative of countries in the developing world, where biomass fuels are also the predominant choice by poor households.

In addition to introducing an extensive list of controls, we pay particular attention to the reverse causality between child mortality and cooking fuel choice. This is important for several reasons. First, a switch to cleaner cooking fuels is a readily available remedial measure subsequent to any prior mortality events that may have been caused by poor air quality in a household.[9] 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).[10]

For identification, therefore, we leverage two instrumental variables for household fuel

[9]There exist vast literature on the household's averting behavior for clean air in response to adverse impacts of outdoor air pollution on health (Gerking and Stanley, 1986; Mansfield et al., 2006; Graff Zivin and Neidell, 2009; Moretti and Neidell, 2011; Barreca et al., 2016; Deschenes et al., 2017; Ito and Zhang, 2020). The studies provide evidence that households react to changes in health outcomes due to air pollution by adjusting their behavior, adopting new technologies, and investing in protective goods in response to health shocks.

[10]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).

choice – the speed of change in forest cover and agricultural land ownership. 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. Our estimate of the causal effect of IAP, as determined by cooking fuel choice on under-five mortality, thus rely 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. Another identification assumption is monotonicity or no-defiers (Imbens and Angrist, 1994), which we check by investigating the cumulative distribution function of polluting fuel choice, separately for those with $Z_i = 1$ and $Z_i = 0$, where Z_i is an instrument. We show that the monotonicity assumption holds as an increase in the instrument leads to a monotonic change in the uptake of the treatment.

Our analysis shows that a household relying on a polluting fuel for cooking has a 4.7 percent higher probability of experiencing under-five child mortality. In addition, we carefully address the issue of heterogeneous treatment effect and explore heterogeneity according to children's age group, as well as household size. A child's age is widely acknowledged as a key factor determining tolerance to environmental hazards (Black et al., 2003; Gurley et al., 2013; Ezeh et al., 2014; Naz et al., 2015, 2016). Household size, particularly in the Indian context, can change the practical feasibility of using stoves for clean fuels, typically appropriate for smaller families, and a child's proximity to cookstoves or exposure to IAP. We find that IAP contributes to infant mortality in households with fewer than ten members, particularly, poor indoor air quality raises the risk of infant mortality for households with 3-6 members. For instance, a household using dirty fuels for cooking has 11.6 and 9.0 percent higher probability of having neonatal mortality incidence for those with 3-4 and 5-6 members, respectively.

The local average treatment effect (LATE) of IAP on infant mortality is highly heterogeneous by household size, and households with five and six members mainly drive our key LATE estimates. Furthermore, we provide additional results on the disproportionate impact of IAP on neonatal mortality. Finally, our findings complement Imelda (2018, 2020), which provide the lower-bound estimates on the health impacts of clean cooking fuels due to the absence of biomass fuels in consideration, but we include biomass fuels in the list of polluting fuels. When only biomass fuels are counted as polluting fuels, the probability of experiencing under-five mortality decreases from 4.7 to 4.0 percent for a household using biomass fuels for cooking.

We assess the robustness of these findings by estimating a variety of specifications with

additional controls and fixed effects. Particularly, we test how our main results change when we take into account globally the second-largest IAP-reducing policy, National Biomass Cookstoves Initiative, and find that the results are strongly robust to the addition of this widely implemented state-level voucher program in India, and the program is not associated with IAP, which is consistent with previous findings of IAP-limitation policy failures by Hanna et al. (2016). By plotting the cumulative distribution functions of polluting fuel use for households (i) with and without agricultural land, and (ii) living in a region with an above- and below-median speed of change in forest cover, we argue in favor of the monotonicity condition applying in our setting since the cumulative distribution functions do not cross. Following Lee (2018), we also correct the standard errors given that we have multiple instruments and LATEs, and our estimates and derived conclusions are robust and qualitatively identical.

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 paper offers the first estimates of the causal effect of IAP, as proxied by cooking fuel choice that includes the use of biomass fuels, on infant mortality while addressing the endogeneity in the relationship between cooking fuel choices and mortality. 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 any empirical settings to date, perhaps due to the challenge in finding valid instrumental variables.[11]

Second, we utilize the NFHS data sets — a widely-accepted gold standard for research in the developing world — covering 601,509 representative households from all 36 states and 640 districts of India over the last 25 years. Our goal is to complement earlier studies by externally validating the link between IAP and infant mortality as shown in the RCTs of Duflo et al. (2008a) and Hanna et al. (2016) in Orissa and Diaz et al. (2007) and Smith-Sivertsen et al. (2009) in the city of San Marcos, for example. A detailed and large-scale data set 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. The data set also allows us to cast a wide net in measuring child mortality, and accordingly to show the nuanced impact of IAP on the mortality risks of four different age-groups including neonatal, post-neonatal, child, and under-five. As an additional improvement over earlier studies, we consider a total of 10 types of cooking fuels to include the choice of biomass fuels as a major

[11] 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.

contributor to IAP that is not yet rigorously addressed in the literature.

Finally, we provide a careful examination of the heterogeneity in the mortality risks implied by IAP. Our analysis covers heterogeneous treatment effects by the child's age and household size. First, we find that switching from clean to polluting fuels leads to an increase in neonatal and under-five mortality, but not post-neonatal and child mortality. The existing economic studies such as Hanna et al. (2016) and Imelda (2018) do not examine the heterogeneous effect of IAP on a child's health by age group, however, are limited to only infants (children with less than one year of age). Imelda (2020) exceptionally investigates the heterogeneity by child's age groups by dividing the infant mortality into perinatal and post-neonatal mortality. Second, we check the heterogeneity of the IAP-mortality link by the number of household members. The results suggest that infants who live in families with fewer than ten members, in particular, those with five and six members, using dirty cooking fuels, are subject to a greater risk of mortality.

The rest of the paper is structured as follows. Section 2 provides the background on IAP and child mortality in India and presents 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 dirty fuels to cook their meals and as the main energy source. This section first discusses 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 μg/m^3 (Smith, 2000). Saksena et al. (1992) finds higher concentrations of PM_{10} (20,000 μg/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), IAP is responsible for 3.5% of the total burden of disease in India (Bonjour et al., 2007), while 20% of deaths among children aged under-five can be attributed to IAP due to polluting fuels use (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 around 2 billion days of work lost due to the diseases caused by IAP in India 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 Initiative (NBCI). 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, its 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.[12] There is some variation in the mortality rate by age group: the neonatal mortality rate (defined as the likelihood of passing away during the first 28 days of early life) is the highest, followed by post-neonatal mortality (measured as the likelihood of passing away between approximately the first month after birth and end of the first year of life) and then child mortality (assessed as the likelihood of passing away between ages of 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 3.0% in 1992 to 1.3% in 2016, and child mortality rate from 1.1% in 1992 to 0.3% in 2016 for those using polluting fuels for cooking.[13]

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 investigate the causal effect of indoor air pollution on infant mortality, we specify the following relationship:

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

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

[12] 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.

[13] 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.

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.[14]

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 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

[14] 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.

PSUs with Census locations at the village-level.[15] 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 National Family Health Survey (NFHS) in India. The NFHS collects individual-level data on mortality incidence and other socio-economic characteristics for every member in the sample household. Additionally, it also 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 a household uses, allowing us to approximate indoor air quality at the household level. To date, four rounds of the survey have been conducted since 1992–93.[16] 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 879,495 ever-married women of reproductive ages between 15–49 years (260,289 from urban and 619,206 from rural areas)

[15] 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.

[16] 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).

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 270,559 singleton live-born children, of whom 18,168 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 are more likely 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 are likely to be large. Our primary outcome variable is under-five mortality. In addition, we consider three preceding age groups, including child, post-neonatal, and neonatal mortality. Neonatal mortality is a death occurred over the first 28 days of life; post-neonatal mortality is a death occurred between one month and the first birthday; child mortality is a death occurred between exact ages one and five.

4.2 Cooking Fuel Choice and Other Controls

The key explanatory variable in our analysis is the cooking fuel type, a proxy for indoor air pollution. Ten types of cooking fuel are reported in the NFHS datasets, and we classify these fuels into two groups, clean and polluting, based on level of smoke produced from cooking. The clean fuels include biogas, liquid petroleum gas (LPG) or natural gas, and electricity while polluting fuels include animal dung, agricultural waste, straw, shrubs or grass, firewood, charcoal, coal or lignite, and kerosene. Note that no household reported using more than one type of fuel for cooking 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),[17] mother's educational attainment (secondary/higher, primary, or no education), type of house (pucca, semi-pucca, or kachha), and number of household members[18] are included as potential socio-economic factors (Bassani et al., 2010; Ezeh et al., 2014; Naz et al., 2015, 2016).

Age of the mother (<20, 20–29, 30–39, and 40–49 years) and gender of the child are also considered as potential determinants of the IAP-infant mortality link. Mother's status of breastfeeding (ever or never breastfed) and place where food is cooked (in the living room, in the kitchen separate from the living room, in a separate building, or outdoors)[19] 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,[20] whereas breastfeeding protects from under-five mortality, particularly in the neonatal and post-natal periods.[21]

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.

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

[18]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 seven members, but there are households with as many as 46 people (maximum is 41 in the NFHS-4 data). 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.

[19]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.

[20]See, for example, Edwards and Langpap (2012); Gurley et al. (2013); Naz et al. (2015, 2016).

[21]See, for example, Cushing et al. (1998); Arifeen et al. (2001); Black et al. (2003); Ezeh et al. (2014).

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.[22] We 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 (5.9% 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

[22] Kuo and Azam (2019) recently attempted to determine the drivers of household's choice of cooking fuel 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 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.

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 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.

Since we essentially estimate a local average treatment effect (LATE) using an IV method, a monotonicity assumption is required to be satisfied (Imbens and Angrist, 1994). This assumption demands, in our context, that IAP does not decrease in a household when the household owns agricultural land or lives in an area with a forest growing at speed above its median, or vice versa. In other words, the monotonicity assumption implies that $D_{1h} - D_{0h} \geq 0$ for all household h, or vice versa, where potential outcomes for IAP (treatment or regressor of interest), $D_h = \{0, 1\}$, against an instrument, $Z_h = \{0, 1\}$, is described as $D_h = D_{0h}(1 - Z_h) + D_{1h}Z_h$. To check whether the monotonicity assumption is valid in our context, we plot the cumulative distribution function (CDF) of polluting fuel choice separately for those with and without agricultural land. For our other instrument, we also investigate the CDF of cooking fuel choice for households at above and below median speed of change in forest cover. The CDFs of polluting fuel use against our two instruments, binary variables, are displayed in Figure 5, and we argue that the monotonicity assumption is valid because the CDFs do not cross implying the first-order stochastic dominance.

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 natural gas, 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 (top panel of Table 3) and gender (bottom panel of Table 3) of the household head.

5 Results

In this section, we first present the estimated average marginal effects[23] 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, 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 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 indoor air pollution (IAP) is linked 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 dummies controls for time-variant spatial factors including state attributes (e.g., characteristics of state magistrate, whether there is any government program regarding the child health service in the state,

[23]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

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

where $f(\cdot)$ is the probability distribution 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 access to medical facilities) and local characteristics (e.g., distance from urban areas and large cities, percentage of districts, sub-districts, 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 provides the estimates for child mortality. The average marginal impact of IAP on child mortality decreases significantly from 0.8 to 0.09 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 and is the age group that spends most of the time with mother. Overall our results offer that the harmful impact of IAP on child 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 immunity and by limiting a child's time spent with mother while cooking.

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; Black et al., 2003; 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* × *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.[24]

[24]The bivariate correlations of under-five mortality with agricultural land ownership and speed of change in forest cover are 0.0036 (SE: 0.0020, p-value: 0.07) and -0.0112 (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 6a. However, Figure 6b 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

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 binary variable, which takes value 1 if household owns land for agricultural purposes in 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 0.048.

Column (1) in top panel 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.[25] Columns (2)–(5) in top panel 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 0.047. In other words, a family relying on polluting fuel for cooking has a 4.7 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

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.

[25] 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, Panel A, indicates that our endogenous regressor not perfectly correlated with the instruments. Hence, we consider that the value of F-statistic reflects our sample size.

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.

The existing economic studies that causally estimate the impact of IAP on infant mortality such as Imelda (2018, 2020) is limited to an impact of kerosene to LPG, and thereby leaving out the dirtiest indoor fuels, the biomass fuels including animal dung, agricultural waste, straw, shrubs, grass, and firewood. However, the literature suggests that (i) a vast majority of households use polluting fuels indoors use biomass fuels, (ii) biomass fuels, being the cheapest, are most popularly used by the poor, and (iii) biomass fuels are far more health-endangering than kerosene, for example (Fullerton et al., 2008). Hence, to check how our results change if we redefine polluting fuels to include only biomass fuels – since these are the key fuel choice that our IVs are supposed to have impact on. The causal effects of biomass fuel for cooking on under-five and neonatal mortality range from 2.9 to 4.0 percent (Columns (2)–(5) in bottom panel of Table 10). These are lower than the counterparts above of 3.4 and 4.7 percent since biomass fuels are part of our complete list of polluting fuels.

Since our local average treatment has heterogeneous effects, the moment condition evaluated at two-stage least squares (2SLS) estimand, which is a positively-weighted average of multiple local average treatment effects (LATEs) given more than one instrument, would be misspecified (Lee, 2018). The conventional heteroskedasticity-robust variance of the 2SLS estimator would be misleading. Thus, Table 10 also reports heteroskedasticity standard errors robust to multiple-LATEs, and our conclusion is qualitatively the same.

Additional heterogeneity results: We make use of data splits to analyze heterogeneous LATEs by household size in addition to heterogeneity by child's age. A subpopulation of households with fewer than ten members is found as the driver of our main estimates of local average effects of polluting fuel use for cooking on infant mortality for four different age-groups presented in the top panel of Table 10, given that such households constitute 92 percent of our sample (top panel of Table 11). Interestingly, a positive effect of IAP on post-neonatal mortality turns out to be statistically significant at the 10% level when the sample is restricted to households with fewer than ten members, with coefficient estimate of 2.4 percent. Regressions in panels B and C of Table 11 using other two subpopulations of households with more than ten members suggest that speed of change in forest cover is not relevant to the polluting fuel choice for cooking, and local average effects of IAP is not statistically significant at the 10 percent level for all age-groups.

We further analyze the heterogeneous treatment effects of IAP on infant mortality by household size, focusing on five subpopulations of households with ten or fewer members: those with fewer than 2 members, 3-4 members, 5-6 members, 7-8 members, and more than 8 members (Figure 7). The joint F-statistic on the excluded instruments indicates that the two IVs introduce meaningful variations in fuel choice in all regressions estimated on subpopulations except for households with one and two members. The results suggest that the local average effect of IAP on under-five mortality is statistically significant only for households with five and six members at 1 percent significance level, with an estimate of 11.1 percent.[26] In addition, a household with 3-4 and 5-6 members using polluting fuels for cooking has 11.6 and 9.0 percent higher probability of experiencing neonatal mortality, respectively. Higher probability of younger child's death in relatively smaller families with three and four household members could be due to a lack of family members who can take care of infants except for a mother cooking using dirty fuels. The effects of IAP on infant mortality for all age-groups are essentially zero for households with 1-2, 7-8, and 9-10 members. Post-neonatal and child mortality are not affected by IAP in any of the households with different pairs of family members. Supplementary Appendix Table S.1 provides the detailed results.[27]

Results above indicate that infants living in a relatively smaller family are subject to a

[26]The local average effects of IAP on under-five and post-neonatal mortality are positive, with coefficients ranging from 0.102 to 0.227, and significantly different from zero for households with three and four members at the 1% level. However, Hansen's J-statistic suggests that the excluded IVs in under-five and post-neonatal mortality regressions for households with three and four members are not exogenous since the null hypothesis of the Sargan-Hansen test is rejected at the 5% and 1% level, respectively.

[27]We also investigate how classification affects our results of heterogeneous LATE by household size by choosing other sets of subpopulations of households with fewer than ten or twelve members. We first combine 3-4 and 5-6 members into one group and 7-8 and 9-10 into another single group. With this new classification, our results suggest that local average treatment effects of IAP on under-five, post-neonatal, and neonatal mortality are positive and statistically significant at 1 percent level. However, the excluded IVs in these regressions are hardly exogenous according to Hansen's J-statistic (Supplementary Appendix Figure S.1 and Table S.2). Supplementary Appendix Figure S.2 and Table S.3 show the heterogeneous treatment effects of IAP on infant mortality when the sample of households with twelve or fewer members is classified into four groups (3-by-3) in terms of household size. The result reveals that IAP increases the neonatal mortality for households with 4-6 members. The local treatment effects of IAP on under-five and post-neonatal mortality are positive and statistically significant at the 1% level; however, the excluded IVs are not exogenous in these two particular regressions according to Hansen's J-statistic. The effects of IAP on infant mortality for all age-groups are not significantly different from zero for households with members other than 4-6. When the sample of households with twelve or fewer members is classified into three groups (4-by-4) in terms of household size, we also find that a household with 1-4 and 5-8 members relying on polluting fuel for cooking respectively has a 14.0 and 5.1 percent higher probability of experiencing neonatal mortality (Supplementary Appendix Figure S.3 and Table S.4). The local average effect of IAP on under-five mortality is 6.3 percent and statistically significant at the 1 percent level. Although the coefficient estimates on IAP in under-five and post-neonatal mortality are positive and statistically significant at the 1% level, the null hypothesis of the Sargan-Hansen test is rejected at the 10% and 5% level, respectively, suggesting that the excluded IVs are endogenous in those two regressions. The heterogeneous treatment effects of IAP on infant mortality for all age-groups are essentially zero for households with 9-12 members.

greater risk of mortality due to IAP, whereas IAP in a relatively larger household has no significant effect on infant mortality. There are several conjectures that can explain why a household with few members has a greater risk of child mortality. For example, it could be because no one is available other than an individual who is cooking to take care of the child and keep them away from the cooking place while food has been cooked. However, one might suggest that a larger household using solid fuel has essentially a zero risk of experiencing child mortality arguing that cooking in a separate building or outdoors is more practical for such a household, while there is more probability of infant mortality in a smaller household because a separate building is not necessary to cook for those few household members, leading such household to cook in the same location as they live in, which means the child is closer to IAP. But household size is negatively correlated with cooking in a separate building (ρ: -0.003, SE: 0.002, p-value: 0.213) and outdoors (ρ: -0.005, SE: 0.002, p-value: 0.020), and thus the latter argument is less likely valid.

We checked if our result of heterogeneous treatment effects by household size is due to the number of household members or location of cooking by adding the location of cooking place to the heterogeneity analysis. In particular, we split the subsamples again by cooking place location and estimate local average effects of IAP on infant mortality for smaller households that cook their food in a separate building and larger households that make their meal in the same room as they live in (unreported). Then we find no clear evidence that location of cooking place drives our household size heterogeneity results. Hence, we argue that a hazardous impact of IAP on infant mortality is present in smaller households using polluting fuels for cooking due to lack of child caretakers because our heterogeneity results above stay the same when restricting the sample of smaller households to those cooking in a separate building and larger household to those cooking in their living room.

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. 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 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 0.539 to 0.569.

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.[28] Hence, we additionally control for states where improved cookstoves program has 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 (Holdren and Smith, 2000). The assigned values to different types of fuels used for cooking are: 1 = electricity, 2 = LPG or natural gas, 3 = biogas, 4 = kerosene, 5 = coal or lignite, 6 = charcoal, 7 = firewood, 8 = straw, shrubs or grass, 9 = agricultural waste, 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 0.8 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 0.8 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

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

in forest cover, we find that the results are also exceptionally robust to the utilization of 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 depend on dirty cooking fuels, and it constitutes the largest source of poor indoor air quality. 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 polluting fuels for cooking or IAP in general 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 145,000 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, including most health-endangering biomass 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 4.7 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 child and post-neonatal mortality. The present study also offers the first empirical analysis showing that the local average effect of IAP on infant mortality is significantly heterogeneous by a

child's age and household size.

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 causal impact 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 puts children at risk. The WHO guidelines for indoor air quality (World Health Organization, 2014) recommend that kerosene be considered as a solid fuel and encourage not to use it indoor. However, nearly 1 billion people who do not have electricity access still use kerosene to light up their homes. Besides creating IAP, kerosene use inside the house also generates other sources of risks such as fires and carbon monoxide poisoning. Therefore, we might have underestimated the causal estimate of the impact 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 has an impact on other socio-economic and health outcomes in addition to infant mortality. 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 household energy choice.

References

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

Arceo, Eva, Rema Hanna, and Paulina Oliva. 2016. "Does the Effect of Pollution on Infant Mortality Differ between Developing and Developed Countries? Evidence from Mexico City." *The Economic Journal*, 126(591): 257–280.

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, RS 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.

Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S. Shapiro. 2016. "Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century." *Journal of Political Economy*, 124(1): 105–159.

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.

Beach, Brian, and W. Walker Hanlon. 2018. "Coal Smoke and Mortality in an Early Industrial Economy." *The Economic Journal*, 128(615): 2652–2675.

Benshaul-Tolonen, Anja. 2019. "Local Industrial Shocks and Infant Mortality." *The Economic Journal*, 129(620): 1561–1592.

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*. Geneva: 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.

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.

Cesur, Resul, Erdal Tekin, and Aydogan Ulker. 2017. "Air Pollution and Infant Mortality: Evidence from the Expansion of Natural Gas Infrastructure." *The Economic Journal*, 127(600): 330–362.

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): 141–69.

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): 151–72.

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.

Deschenes, Olivier, Michael Greenstone, and Joseph S. Shapiro. 2017. "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program." *American Economic Review*, 107(10): 2958–2989.

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.

Duflo, Esther, Michael Greenstone, and Rema Hanna. 2008a. "Cooking Stoves, Indoor Air Pollution and Respiratory Health in Rural Orissa." *Economic and Political*

Weekly, 43(32): 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.

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.

Fullerton, Duncan G., Nigel Bruce, and Stephen B. Gordon. 2008. "Indoor Air Pollution from Biomass Fuel Smoke is a Major Health Concern in the Developing World." *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(9): 843–851.

Gerking, Shelby, and Linda R. Stanley. 1986. "An Economic Analysis of Air Pollution and Health: The Case of St. Louis." *The Review of Economics and Statistics*, 68(1): 115–121.

Graff Zivin, Joshua, and Matthew Neidell. 2009. "Days of Haze: Environmental Information Disclosure and Intertemporal Avoidance Behavior." *Journal of Environmental Economics and Management*, 58(2): 119–128.

Graff Zivin, Joshua, and Matthew Neidell. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review*, 102(7): 3652–73.

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. eds. by Sander Greenland Kenneth J. Rothman, and Timothy L. Lash, Philadelphia: 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.

Holdren, John P., and Kirk R. Smith. 2000. "Energy, the Environment, and Health." In *The World Energy Assessment: Energy and the Challenge of Sustainability*. ed. by José Goldemberg, New York: UN Development Programme, 61–110.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–475.

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

Imelda. 2020. "Cooking that Kills: Cleaner Energy Access, Indoor Air Pollution, and Health." *Journal of Development Economics*, 147.

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.

Ito, Koichiro, and Shuang Zhang. 2020. "Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China." *Journal of Political Economy*, 128(5): 1627–1672.

Jayachandran, Seema. 2009. "Air Quality and Early-life Mortality Evidence from Indonesia's Wildfires." *Journal of Human Resources*, 44(4): 916–954.

Jia, Ruixue, and Hyejin Ku. 2019. "Is China's Pollution the Culprit for the Choking of

South Korea? Evidence from the Asian Dust." *The Economic Journal*, 129(624): 3154–3188.

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.

Lee, Seojeong. 2018. "A Consistent Variance Estimator for 2SLS When Instruments Identify Different LATEs." *Journal of Business & Economic Statistics*, 36(3): 400–410.

Liu, Ziming, Jia Li, Jens Rommel, and Shuyi Feng. 2020. "Health Impacts of Cooking Fuel Choice in Rural China." *Energy Economics*, 89: 104811.

Mansfield, Carol, F. Reed Johnson, and George Van Houtven. 2006. "The Missing Piece: Valuing Averting Behavior for Children's Ozone Exposures." *Resource and Energy Economics*, 28(3): 215–228.

Moretti, Enrico, and Matthew Neidell. 2011. "Pollution, Health, and Avoidance Behavior Evidence from the Ports of Los Angeles." *Journal of Human Resources*, 46(1): 154–175.

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.

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, Stanford: Stanford Institute for Theoretical Economics.

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

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.

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.

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

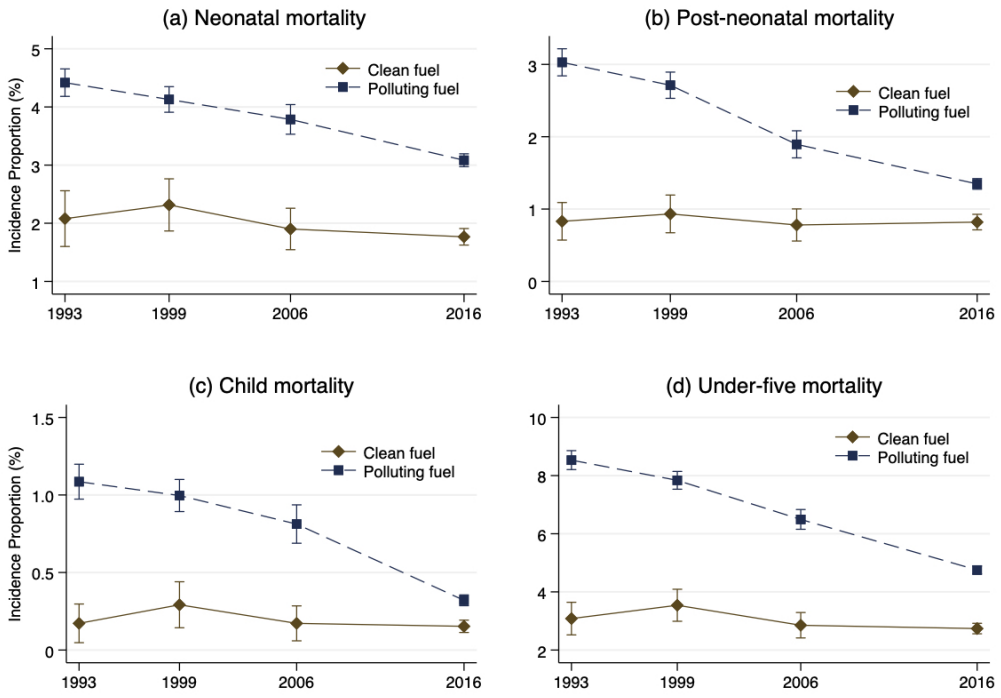
World Health Organization. 2016. "Burning Opportunity: Clean Household Energy for Health, Sustainable Development, and Wellbeing of Women and Children."

World Health Organization. 2018a. "Household Air Pollution and Health, Fact Sheet No. 292."

World Health Organization. 2018b. "World Health Statistics 2018: Monitoring Health for the SDGs."

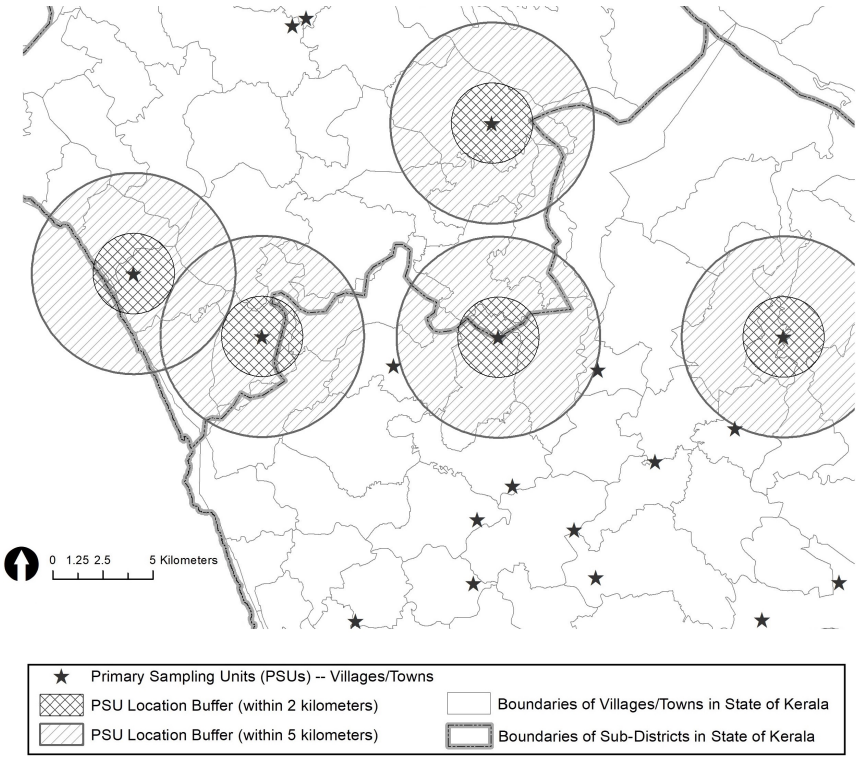
Zhang, Junfeng, and Kirk R. Smith. 2007. "Household Air Pollution from Coal and Biomass Fuels in China: Measurements, Health Impacts, and Interventions." *Environmental Health Perspectives*, 115(6): 848–855.

Figure 1: Mortality Trend in Different Child's Age-Groups by Fuel Type in India



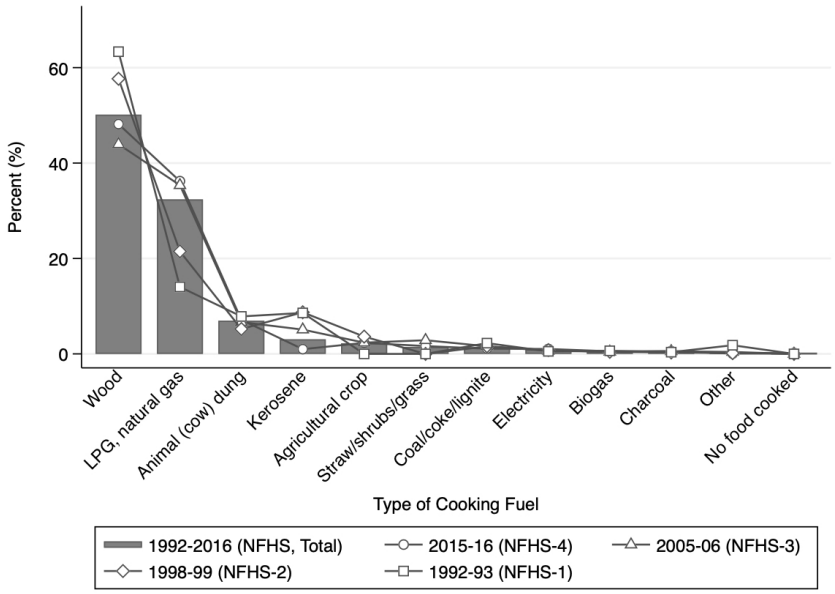
Notes: Based on NFHS datasets 1992–93, 1998–99, 2005–06, and 2015–16. In the medical literature, the measure of incidence proportion (or cumulative incidence) is described as the fraction of children alive at the start of a period who die over that period (Greenland and Rothman, 2008; Centers for Disease Control and Prevention, 2006). To adjust for multiple-stage cluster sample 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)



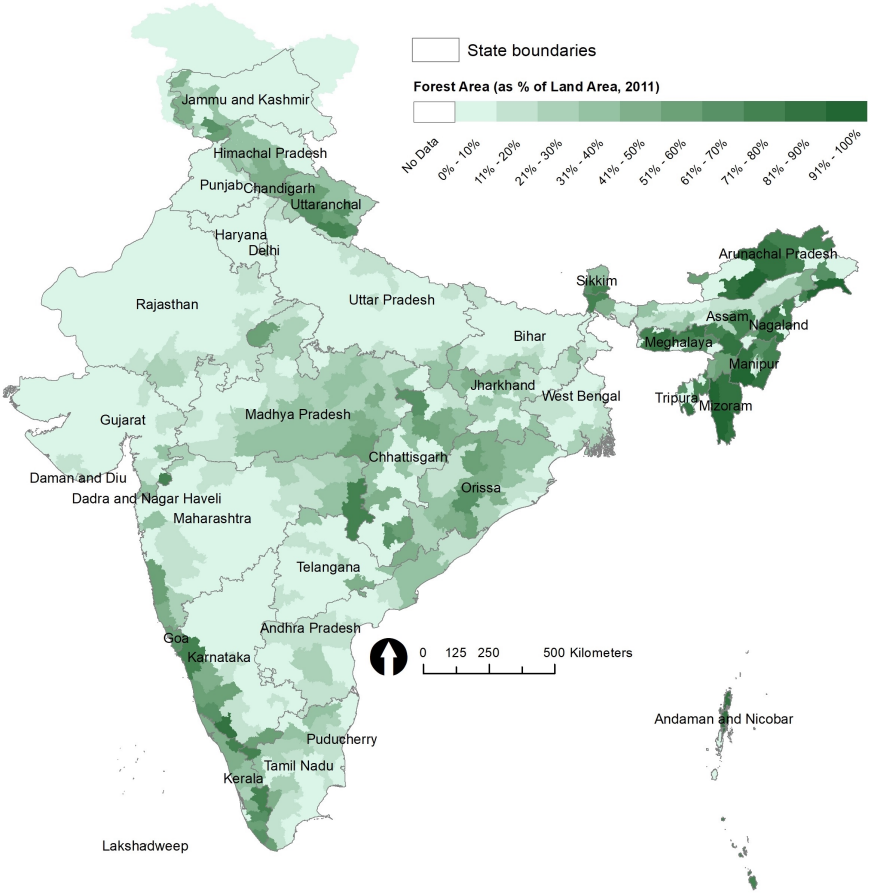
Notes: The figure shows how the PSU points are displaced in NFHS-4 (2015–16) survey based on few PSU points in Kerala district. In order to ensure that respondent confidentiality is maintained, the GPS (latitude/longitude positions) of respondent locations are randomly displaced according to the "random 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.

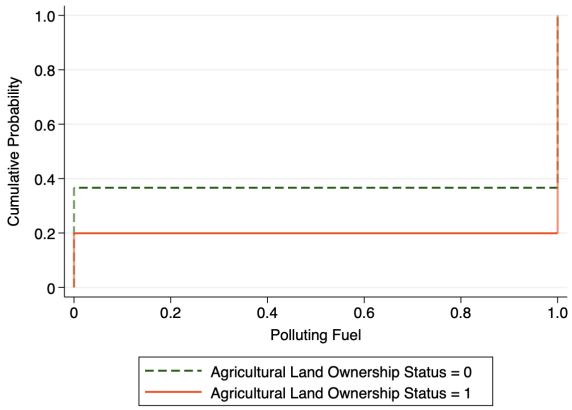
Figure 4: India's District-Wise and Satellite-based Forest Cover



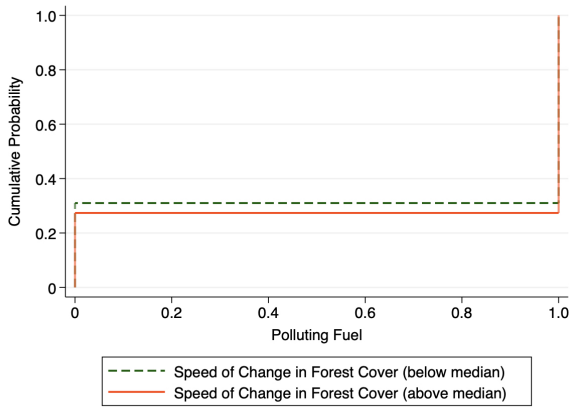
Notes: Based on a satellite-based data on forest cover from the Planning Commission of India. The figure depicts the 2011 district-wise forest cover (measured by percentage of geographical area covered by forests) in India. The forest cover includes all types of forests (different canopy density classes) 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: Cumulative Distribution Functions (CDFs) of Polluting Fuel Choice

(a) By Ownership Status of Agricultural Land



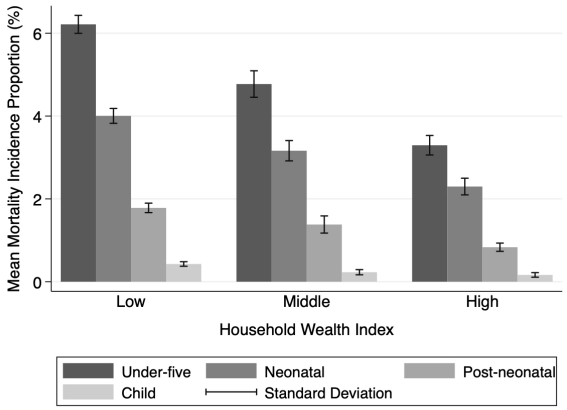
(b) By Speed of Change in Forest Cover



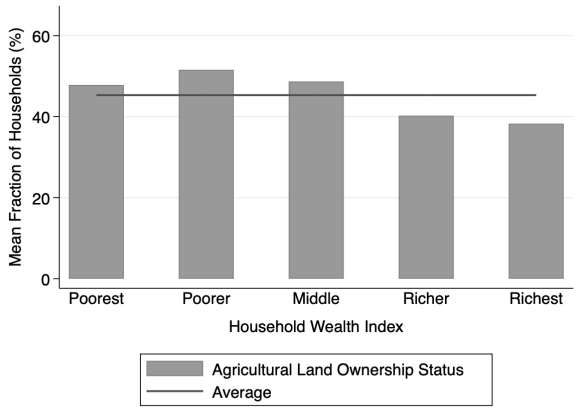
Notes: Based on NFHS-4 (2015–16) dataset. This figure plots the cumulative distribution function of polluting fuel choice for households with and without land for agricultural purposes (left panel) and those living in an area with forest cover growing at speed above and below the median speed of change in forest cover (right panel).

Figure 6: Mortality in Different Age-Groups of Children and Ownership Status of Agricultural Land by Household Wealth in India

(a) Infant Mortality



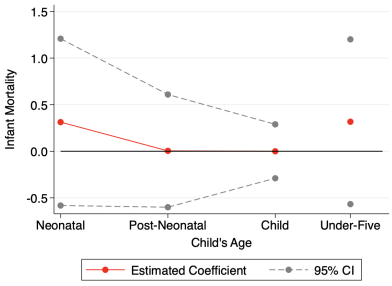
(b) Ownership of Agricultural Land



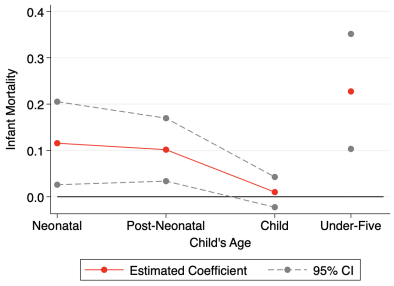
Notes: Based on NFHS datasets 1992–93, 1998-99, 2005–06, and 2015–16. Panel (a) shows that under-five mortality incidence proportion is higher in households with lower wealth, suggesting the probability of mortality decreases as a family becomes wealthier. The "low wealth" is the bottom 40% of households, "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.

Figure 7: Heterogeneous Effects of IAP on Infant Mortality by Child's Age and Household Size using IV (2SLS) Regressions for Households with Fewer than Ten Members (Classified as 2-by-2 groups) (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

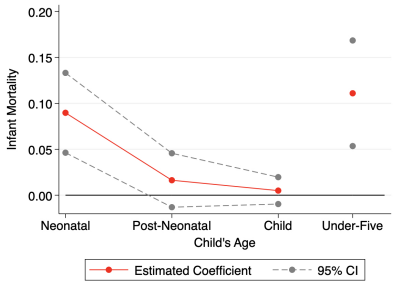
(a) Household size = $\{1, 2\}$, N = 1,239



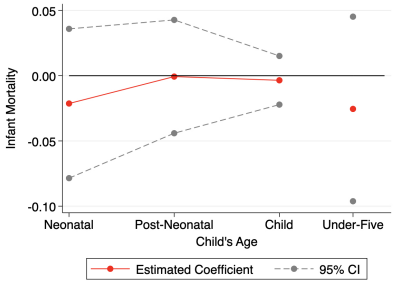
(b) Household size = $\{3, 4\}$, N = 48,162



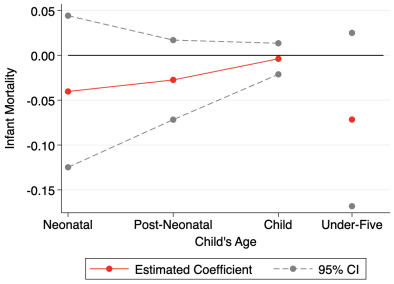
(c) Household size = $\{5, 6\}$, N = 70,990



(d) Household size = $\{7, 8\}$, N = 39,945



(e) Household size = $\{9, 10\}$, N = 19,809



Notes: The figure presents heterogeneous treatment effects of IAP (*defined by polluting fuel use*) on infant mortality by both child's age and household size using five distinct subpopulations of households with fewer than ten members (based on our findings in Table 11) covered in NFHS-4 data. Panels (a)–(e) considers each of the five subsamples from pair of 1-2 to 9-10 members in an orderly fashion. Each panel reports results from the estimation of Equation (1) using IV regression with different dependent variables and similar specifications where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The outcome variable in each IV regression is a binary variable of infant mortality for each of the four different age-groups, and the endogenous regressor or outcome variable in the first-stage regression is whether fuel choice: polluting fuel. The speed of change in forest cover and agricultural land ownership status are used as instruments, and first-stage coefficient estimates on the IVs are both positive and statistically significant at least at the 10% level except for the speed of change in forest cover in the first-stage regression for households with nine and ten members. The F-test on IVs verifies that the instruments generate a plausible variation in polluting fuel for cooking, except for a subpopulation of households with one and two members. The calculation of Hansen's J-statistic is not available for IV regressions of households with one and two members due to a lack of observations. The Hansen's J-statistic suggests that the excluded IVs are not exogenous in under-five and post-neonatal mortality regressions for households with three and four members since a rejection of the null hypothesis of the Sargan-Hansen test is encountered at the 5% and 1% level, respectively. All specifications contain a vector of demographic controls and a 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. Unit of observation: child. Heteroskedasticity-consistent standard errors are clustered by districts.

Table 1: Summary Statistics

	Mean	S.D.	Min	Max	Obs.
Infant mortality (% total live births)					
Under-five	0.053	0.223	0.000	1.000	369,416
Child	0.005	0.069	0.000	1.000	369,416
Post-neonatal	0.017	0.130	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 after birth, Post-neonatal = period between approximately the first month after birth and end of the first year of life, and Child = period from age of one to five. 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		Obs.
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	
Uttar Pradesh	0.074	0.262	0.006	0.080	0.024	0.153	0.044	0.204	56,090
Madhya Pradesh	0.066	0.248	0.007	0.085	0.021	0.144	0.038	0.190	32,007
Odisha	0.059	0.236	0.004	0.066	0.021	0.144	0.034	0.181	16,192
Rajasthan	0.059	0.235	0.005	0.070	0.021	0.144	0.033	0.177	25,435
Bihar	0.059	0.236	0.005	0.071	0.017	0.129	0.037	0.190	33,093
Assam	0.058	0.234	0.007	0.082	0.018	0.134	0.033	0.179	14,393
Chhattisgarh	0.058	0.234	0.005	0.070	0.014	0.116	0.040	0.196	10,695
Andhra Pradesh	0.055	0.228	0.003	0.057	0.018	0.134	0.033	0.179	5,515
Meghalaya	0.050	0.218	0.006	0.075	0.019	0.138	0.025	0.156	6,261
Gujarat	0.050	0.218	0.006	0.078	0.015	0.121	0.029	0.168	12,077
All States/UTs	0.053	0.223	0.005	0.069	0.017	0.130	0.031	0.173	369,416

Panel B. Type of Cooking Fuel (fraction)

States	Mean		S.D.	Obs.
	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		Obs.	Mean		S.D.	Obs.
	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	0.006	0.076	0.046	0.210	0.048	0.215	518	0.931	0.069	0.253	494
Age 20-29	0.058	0.234	0.004	0.066	0.018	0.134	0.036	0.186	62,807	0.812	0.188	0.391	62,067
Age 30-39	0.052	0.222	0.006	0.078	0.018	0.134	0.028	0.164	111,513	0.751	0.249	0.432	110,103
Age 40-49	0.060	0.238	0.007	0.081	0.020	0.139	0.034	0.182	53,663	0.791	0.209	0.407	50,169
Age 50-59	0.049	0.216	0.003	0.056	0.015	0.122	0.031	0.173	60,828	0.732	0.268	0.443	55,654
Age 60-69	0.046	0.209	0.003	0.055	0.014	0.116	0.029	0.168	56,211	0.720	0.280	0.449	52,830
Age 70-79	0.047	0.213	0.004	0.063	0.014	0.119	0.029	0.168	18,920	0.737	0.263	0.440	18,112
Age 80-89	0.052	0.221	0.004	0.063	0.018	0.133	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	0.223	0.005	0.069	0.017	0.130	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	0.005	0.070	0.017	0.130	0.031	0.174	330,884	0.762	0.238	0.426	317,867
Female	0.048	0.214	0.003	0.058	0.016	0.126	0.028	0.166	38,530	0.738	0.262	0.440	36,292
Total	0.053	0.223	0.005	0.069	0.017	0.130	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 Indoor Air Pollution 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		0.011***	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		0.018***	0.018***
		(0.003)	(0.003)
Mother's age: 20-29		-0.008***	-0.008***
		(0.002)	(0.002)
Mother's age: 30-39		-0.012***	-0.012***
		(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**	-0.002**
		(0.001)	(0.001)
Food cooked: In a separate building		-0.002	-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 fixed effects	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
State-by-Year fixed effects	No	No	Yes
Observations	354,161	230,091	230,091
Probit log-likelihood	-71,277	-37,008	-37,000

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 calculated by applying the Delta method. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 5: Probit: The Marginal Impact of Indoor Air Pollution on Child Mortality

	Dependent variable: Child mortality		
	(1)	(2)	(3)
Polluting fuel for cooking	0.0040***	0.0009**	0.0009**
	(0.0004)	(0.0004)	(0.0004)
Place of residence: Rural		0.0004	0.0004
		(0.0004)	(0.0004)
Household wealth: Middle		0.0008	0.0008
		(0.0006)	(0.0006)
Household wealth: Low		0.0015**	0.0014**
		(0.0007)	(0.0007)
Number of household members		-0.0005***	-0.0005***
		(0.0001)	(0.0001)
Mother's age: <20		-0.0032***	-0.0031***
		(0.0011)	(0.0011)
Mother's age: 20-29		-0.0021***	-0.0021***
		(0.0005)	(0.0005)
Mother's age: 30-39		-0.0010**	-0.0010**
		(0.0005)	(0.0005)
Mother's education: Primary		0.0010***	0.0010***
		(0.0004)	(0.0004)
Mother's education: No education		0.0018***	0.0019***
		(0.0003)	(0.0003)
Gender of child: Male		-0.0004*	-0.0004*
		(0.0002)	(0.0002)
Never breastfed		0.0023***	0.0023***
		(0.0002)	(0.0002)
Food cooked: In separate kitchen inside		-0.0002	-0.0002
		(0.0003)	(0.0003)
Food cooked: In a separate building		-0.0002	-0.0002
		(0.0004)	(0.0004)
Food cooked: Outdoors		-0.0001	-0.0002
		(0.0004)	(0.0004)
House type: Semi-pucca		0.0001	0.0001
		(0.0003)	(0.0003)
House type: Kachha		0.0003	0.0003
		(0.0004)	(0.0004)
Year fixed effects	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
State-by-Year fixed effects	No	No	Yes
Observations	351,822	228,039	228,039
Probit log-likelihood	-10,428	-4,074	-4,074

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 calculated by applying the Delta method. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 6: Probit: The Marginal Impact of Indoor Air Pollution on Post-Neonatal Mortality

	Dependent variable: Post-neonatal mortality		
	(1)	(2)	(3)
Polluting fuel for cooking	0.009***	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		-0.004***	-0.004***
		(0.001)	(0.001)
Mother's age: 30-39		-0.004***	-0.004***
		(0.001)	(0.001)
Mother's education: Primary		0.003***	0.003***
		(0.001)	(0.001)
Mother's education: No education		0.005***	0.005***
		(0.001)	(0.001)
Gender of child: Male		-0.001*	-0.001*
		(0.000)	(0.000)
Never breastfed		0.015***	0.015***
		(0.001)	(0.001)
Food cooked: In separate kitchen inside		-0.001	-0.001
		(0.001)	(0.001)
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
		(0.001)	(0.001)
House type: Kachha		0.000	0.000
		(0.001)	(0.001)
Year fixed effects	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
State-by-Year fixed effects	No	No	Yes
Observations	354,161	229,696	229,696
Probit log-likelihood	-29,912	-14,389	-14,389

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 calculated by applying the Delta method. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 7: Probit: The Marginal Impact of Indoor Air Pollution 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 fixed effects	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
State-by-Year fixed effects	No	No	Yes
Observations	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 calculated by applying 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***	1.28***	1.25***	1.25***
		(0.065)	(0.117)	(0.054)	(0.047)
Marginal Effect		0.014***	0.010***	0.013***	0.009***
		(0.003)	(0.004)	(0.003)	(0.001)
Observations	138,063	150,845	32,503	181,791	230,091
	Dependent variable: Child mortality				
Odds Ratio	1.42**	1.55**	0.75	1.24	1.38**
		(0.291)	(0.250)	(0.194)	(0.208)
Marginal Effect		0.004**	-0.001	0.002	0.001**
		(0.002)	(0.001)	(0.001)	(0.000)
Observations	138,063	150,845	31,810	181,791	228,039
	Dependent variable: Post-neonatal mortality				
Odds Ratio	1.42***	1.40***	1.08	1.28***	1.11
		(0.127)	(0.177)	(0.099)	(0.072)
Marginal Effect		0.008	0.001	0.005***	0.001
		(0.005)	(0.002)	(0.002)	(0.001)
Observations	138,063	150,845	32,503	181,791	229,696
	Dependent variable: Neonatal mortality				
Odds Ratio	1.23***	1.18**	1.46***	1.25***	1.30***
		(0.076)	(0.168)	(0.068)	(0.061)
Marginal Effect		0.006**	0.010***	0.007***	0.007***
		(0.002)	(0.003)	(0.002)	(0.001)
Observations	138,063	150,845	32,503	181,791	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, house type, location of cooking place, household wealth, breastfeeding status, gender of child, mother's working status, mother's educational attainment, mother's age, place of residence, 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 calculated by applying 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)

	1st stage	2nd stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting Fuel Use	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.048***	0.0004	0.011	0.037***
		(0.017)	(0.0044)	(0.010)	(0.014)
Owns agricultural land	0.057***				
	(0.002)				
Observations	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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. OLS regression does not drop the state FEs that perfectly explain the child and post-natal 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 IV (2SLS) Regressions (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

	1st stage	2nd stage			
	(1) Polluting/Biomass Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Panel A. Indoor air pollution = Polluting fuel					
Polluting fuel for cooking		0.047**[***]	-0.001	0.014	0.034**[**]
		(0.020)	(0.005)	(0.011)	(0.015)
		[0.018]	[0.005]	[0.010]	[0.014]
Speed of change in forest cover	0.030***				
	(0.011)				
Owns agricultural land	0.054***				
	(0.003)				
Observations	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		0.95	0.78	0.20	1.80
χ^2 p-value		0.33	0.38	0.65	0.18
Panel B. Indoor air pollution = Biomass fuel					
Biomass fuels for cooking		0.040**[***]	-0.001	0.012	0.029**[**]
		(0.017)	(0.004)	(0.009)	(0.013)
		[0.015]	[0.004]	[0.008]	[0.012]
Speed of change in forest cover	0.035***				
	(0.013)				
Owns agricultural land	0.064***				
	(0.003)				
Observations	196,344	196,344	196,344	196,344	196,344
R^2	0.52	0.02	0.00	0.01	0.02
F-stat on IVs	212.02				
Hansen's J-statistic		0.93	0.78	0.19	1.80
χ^2 p-value		0.34	0.38	0.66	0.18

Notes: The first column 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: polluting fuel (top panel) and biomass fuel (bottom panel). 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 (top panel) and biomass (bottom panel) fuels for cooking. Columns (2), (3), (4) and (5) 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 (top panel) or biomass fuel (bottom panel) from the first-stage estimation. The Hansen's J-statistic suggests that the excluded IVs are exogenous. 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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. Unit of observation: child. Heteroskedasticity-consistent standard errors clustered by districts are in parentheses. Robust to multiple-LATEs and heteroscedasticity standard errors (Lee, 2018) of the key regressor and statistical significance based on them are in square brackets. The statistical significances of the key regressors are the same for all regressions when the standard errors are clustered by PSUs. Significance: $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 11: Heterogeneous Effects of Polluting Fuel Use on Infant Mortality by Household Size from IV (2SLS) Regressions (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

	1st stage	2nd stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting Fuel Use	Under-Five	Child	Post-Neonatal	Neonatal
Panel A. Number of household members $\in [1, 10]$					
Polluting fuel for cooking		0.076***	0.002	0.024*	0.050***
		(0.023)	(0.006)	(0.013)	(0.017)
Speed of change in forest cover	0.032***				
	(0.011)				
Owns agricultural land	0.051***				
	(0.003)				
Observations	180,145	180,145	180,145	180,145	180,145
R^2	0.54	0.02	0.00	0.01	0.01
F-stat on IVs	148.80				
Hansen's J-statistic		1.91	0.61	0.66	2.59
χ^2 p-value		0.17	0.43	0.42	0.11
Panel B. Number of household members $\in [11, 20]$					
Polluting fuel for cooking		0.016	0.000	-0.010	0.026
		(0.045)	(0.010)	(0.019)	(0.036)
Speed of change in forest cover	-0.016				
	(0.018)				
Owns agricultural land	0.084***				
	(0.010)				
Observations	15,685	15,685	15,685	15,685	15,685
R^2	0.51	0.01	0.00	0.01	0.01
F-stat on IVs	38.00				
Hansen's J-statistic		0.67	1.00	0.00	0.60
χ^2 p-value		0.41	0.32	0.99	0.44
Panel C. Number of household members $\in [21, 30]$					
Polluting fuel for cooking		-0.076	-0.039	-0.014	-0.024
		(0.072)	(0.031)	(0.015)	(0.048)
Speed of change in forest cover	-0.068				
	(0.102)				
Owns agricultural land	0.264***				
	(0.073)				
Observations	481	481	481	481	481
R^2	0.53	0.04	-0.02	0.11	0.06
F-stat on IVs	6.79				
Hansen's J-statistic		NA	NA	NA	NA
χ^2 p-value		NA	NA	NA	NA

Notes: Based on three subsamples of NFHS-4 data: households with 1-10 (panel A), 11-20 (panel B), and 21-30 (panel C) members. Given that only 33 children are in households with more than 30 members, we ignore such households without loss of any inference. The outcome variable in the first-stage regression is whether fuel choice: polluting fuel. The F-test on the IVs verifies that the instruments generate a plausible variation in polluting fuels for cooking (panels A and B) except for households with 21-30 members (panel C). The Hansen's J-statistic suggests that the excluded IVs are exogenous for IV regressions in panels A and B; however, the statistic is not available in panel C due to a small number of observations. All specifications contain an unreported vector of demographic controls and a 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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. Unit of observation: child. Heteroskedasticity-consistent standard errors clustered by districts are in parentheses. Significance: $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Appendix

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

	1st stage		2nd stage			
	(1) Polluting Fuel Use		(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Polluting fuel for cooking			0.539***	-0.112	0.457	0.569**
			(0.208)	(0.577)	(0.322)	(0.237)
Speed of change in forest cover	0.030***					
	(0.004)					
Owns agricultural land	0.054***					
	(0.002)					
Observations	196,344		196,344	194,831	195,949	196,344
R^2	0.54					
F-stat on IVs	339.59					
Model Wald χ^2			3,883.75	389.81	1,489.68	2,503.41
Model degrees of freedom		50.00	49.00	45.00	48.00	49.00
Model Wald p-value		0.00	0.00	0.00	0.00	0.00
Exogeneity test Wald p-value			0.04	0.73	0.21	0.05
Wald χ^2 test of exogeneity			4.42	0.12	1.56	3.78

Notes: The first column 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) report coefficient estimates from the estimation of Equation (1) using IV probit regression with different dependent variables and a similar specification. 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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled 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 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)

	1st stage	2nd stage			
	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Polluting fuel for cooking		0.047**	-0.001	0.014	0.034**
		(0.019)	(0.005)	(0.010)	(0.015)
Cookstoves Program States (NBCI)	0.056	0.055***	0.004***	0.012***	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)				
Observations	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		1.25	0.67	0.25	1.67
χ^2 p-value		0.26	0.41	0.62	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 educational attainment; and infant 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 dummies are also controlled in every specification. The Hansen's J-statistic suggests that the excluded IVs are exogenous. Unit of observation: 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)

	1st stage	2nd stage			
	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Dirtiness level of cooking fuels		0.008***	-0.0002	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)				
Observations	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		0.28	0.60	0.03	0.61
χ^2 p-value		0.60	0.44	0.87	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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. The Hansen's J-statistic suggests that the excluded IVs are exogenous. Unit of observation: 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)

	1st stage	2nd stage			
	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Polluting fuel for cooking		0.045**	-0.0001	0.012	0.032**
		(0.018)	(0.005)	(0.010)	(0.014)
Forest cover (*satellite-based, 2011*)	0.050***				
	(0.008)				
Owns agricultural land	0.057***				
	(0.002)				
Observations	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		1.28	1.06	0.37	1.70
χ^2 p-value		0.26	0.30	0.54	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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. The Hansen's J-statistic suggests that the excluded IVs are exogenous. Unit of observation: 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)

	1st stage	2nd stage			
	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Polluting fuel for cooking		0.033*	-0.001	0.001	0.033**
		(0.018)	(0.004)	(0.009)	(0.014)
Forest cover (*census-based, 2011*)	0.037***				
	(0.009)				
Owns agricultural land	0.057***				
	(0.002)				
Observations	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		2.40	0.90	3.18	1.16
χ^2 p-value		0.12	0.34	0.07	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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. The Hansen's J-statistic suggests that the excluded IVs are exogenous. Unit of observation: child. Parentheses contain standard errors clustered by PSUs. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Supplementary Appendix

Table S.1: Heterogeneous Effects of IAP on Infant Mortality by Child's Age and Household Size using IV (2SLS) Regressions for Households with Fewer than Ten Members (2-by-2 Classification) (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

	1st stage	2nd stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting Fuel Use	Under-Five	Child	Post-Neonatal	Neonatal
Panel A. Number of household members $= \{1, 2\}$					
Polluting fuel for cooking		0.317	-0.000	0.004	0.313
		(0.451)	(0.148)	(0.308)	(0.457)
Speed of change in forest cover	0.095**				
	(0.039)				
Owns agricultural land	0.043*				
	(0.022)				
Observations	1,239	1,239	1,239	1,239	1,239
R^2	0.53	0.30	0.04	0.08	0.19
F-stat on IVs	5.12				
Hansen's J-statistic		NA	NA	NA	NA
χ^2 p-value		NA	NA	NA	NA
Panel B. Number of household members $= \{3, 4\}$					
Polluting fuel for cooking		0.227***	0.010	0.102***	0.116**
		(0.063)	(0.017)	(0.035)	(0.046)
Speed of change in forest cover	0.038***				
	(0.012)				
Owns agricultural land	0.036***				
	(0.004)				
Observations	48,162	48,162	48,162	48,162	48,162
R^2	0.58	-0.04	0.00	-0.04	-0.01
F-stat on IVs	39.10				
Hansen's J-statistic		4.16	1.41	6.12	2.48
χ^2 p-value		0.04	0.24	0.01	0.11
Panel C. Number of household members $= \{5, 6\}$					
Polluting fuel for cooking		0.111***	0.005	0.016	0.090***
		(0.029)	(0.007)	(0.015)	(0.022)
Speed of change in forest cover	0.036***				
	(0.011)				
Owns agricultural land	0.055***				
	(0.004)				
Observations	70,990	70,990	70,990	70,990	70,990
R^2	0.54	-0.01	0.00	0.01	-0.02
F-stat on IVs	106.63				
Hansen's J-statistic		2.84	0.14	1.99	1.63
χ^2 p-value		0.09	0.71	0.16	0.20

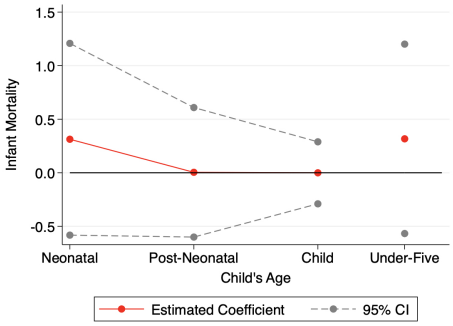
Table S.1: (Continued)

	1st stage	2nd stage			
	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Panel D. Number of household members $= \{7, 8\}$					
Polluting fuel for cooking		-0.025	-0.004	-0.001	-0.021
		(0.036)	(0.010)	(0.022)	(0.029)
Speed of change in forest cover	0.027**				
	(0.013)				
Owns agricultural land	0.052***				
	(0.005)				
Observations	39,945	39,945	39,945	39,945	39,945
R^2	0.52	0.02	0.00	0.01	0.01
F-stat on IVs	66.25				
Hansen's J-statistic		0.04	0.20	1.05	1.30
χ^2 p-value		0.84	0.65	0.30	0.25
Panel E. Number of household members $= \{9, 10\}$					
Polluting fuel for cooking		-0.072	-0.004	-0.027	-0.040
		(0.049)	(0.009)	(0.023)	(0.043)
Speed of change in forest cover	0.016				
	(0.017)				
Owns agricultural land	0.064***				
	(0.008)				
Observations	19,809	19,809	19,809	19,809	19,809
R^2	0.54	0.00	0.00	0.00	0.00
F-stat on IVs	35.11				
Hansen's J-statistic		0.00	0.10	1.86	0.60
χ^2 p-value		1.00	0.75	0.17	0.44

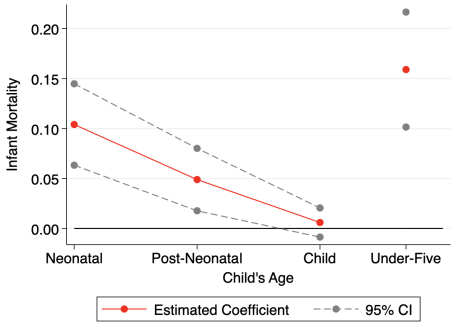
Notes: The table presents heterogeneous treatment effects of IAP on infant mortality by both child's age and household size using five subpopulations of households with fewer than ten members (based on our findings in Table 11) covered in NFHS-4 data. Panel A–E considers each of the five subsamples from pair of 1-2 to 9-10 members in an orderly fashion. The first column provides results from the first-stage regressions of the IV (2SLS) regressions, where the dependent variable is a binary variable of whether fuel choice: polluting fuel. Columns (2), (3), (4) and (5) report results from the estimation of Equation (1) using IV regression with different dependent variables and similar specifications where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The outcome variable in an IV regression is a binary variable of infant mortality for each of the four different age-groups. All specifications contain an unreported vector of demographic controls and a 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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. Unit of observation: child. Heteroskedasticity-consistent standard errors clustered by districts are in parentheses. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Figure S.1: Heterogeneous Effects of IAP on Infant Mortality by Child's Age and Household Size using IV (2SLS) Regressions for Households with Fewer than Ten Members (2-by-2 Classification, Groups are Combined) (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

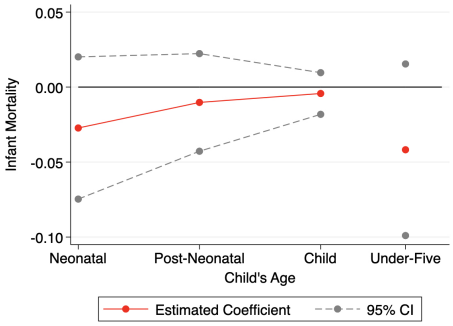
(a) Household size = $\{1, 2\}$, N = 1,239



(b) Household size = $[3, 6]$, N = 119,152



(c) Household size = $[7, 10]$, N = 59,754



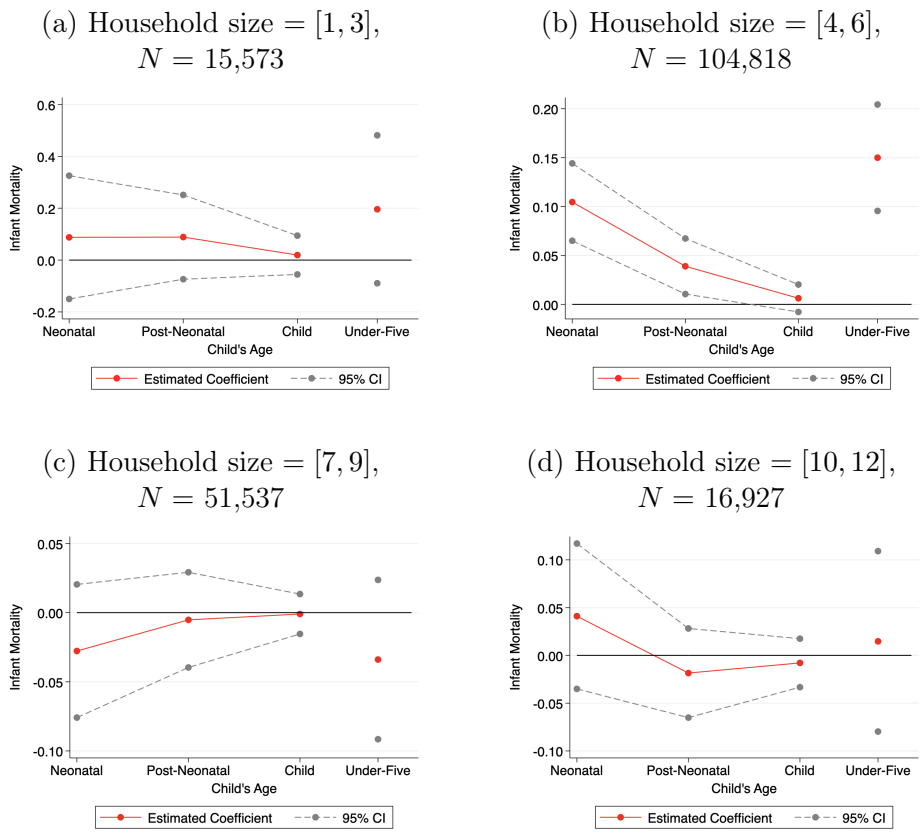
Notes: The figure presents heterogeneous treatment effects of IAP (*defined by polluting fuel use*) on infant mortality by both child's age and household size using three distinct subpopulations of households with fewer than ten members (based on our findings in Table 11) covered in NFHS-4 data. Panels (a)–(c) considers each of the three subsamples from 1-2, 3-6, and 7-10 members in an orderly fashion. Each panel reports results from the estimation of Equation (1) using IV regression with different dependent variables and similar specifications where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The outcome variable in each IV regression is a binary variable of infant mortality for each of the four different age-groups, and the endogenous regressor (i.e., the outcome variable in the first-stage regression) is whether fuel choice: polluting fuel. The speed of change in forest cover and agricultural land ownership status are used as instruments, and first-stage coefficient estimates on the IVs are both positive and statistically significant at least at the 10% level. The F-test on IVs verifies that the instruments generate a plausible variation in polluting fuel for cooking, except for a subpopulation of households with one and two members. The calculation of Hansen's J-statistic is not available for IV regressions of households with one and two members due to a lack of observations. The Hansen's J-statistic suggests that the excluded IVs are not exogenous in under-five, post-neonatal and neonatal mortality regressions for households with 3-6 members since a rejection of the null hypothesis of the Sargan-Hansen test is encountered at least at the 5% level. All specifications contain a vector of demographic controls and a 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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. Unit of observation: child. Heteroskedasticity-consistent standard errors are clustered by districts.

Table S.2: Heterogeneous Effects of IAP on Infant Mortality by Child's Age and Household Size using IV (2SLS) Regressions for Households with Fewer than Ten Members (2-by-2 Classification, Groups are Combined) (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

	1st stage	2nd stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting Fuel Use	Under-Five	Child	Post-Neonatal	Neonatal
Panel A. Number of household members $= \{1, 2\}$					
Polluting fuel for cooking		0.317	-0.000	0.004	0.313
		(0.451)	(0.148)	(0.308)	(0.457)
Speed of change in forest cover	0.095**				
	(0.039)				
Owns agricultural land	0.043*				
	(0.022)				
Observations	1,239	1,239	1,239	1,239	1,239
R^2	0.53	0.30	0.04	0.08	0.19
F-stat on IVs	5.12				
Hansen's J-statistic		NA	NA	NA	NA
χ^2 p-value		NA	NA	NA	NA
Panel B. Number of household members $= [3, 6]$					
Polluting fuel for cooking		0.159***	0.006	0.049***	0.104***
		(0.029)	(0.007)	(0.016)	(0.021)
Speed of change in forest cover	0.037***				
	(0.011)				
Owns agricultural land	0.048***				
	(0.003)				
Observations	119,152	119,152	119,152	119,152	119,152
R^2	0.56	-0.02	0.00	-0.01	-0.02
F-stat on IVs	112.01				
Hansen's J-statistic		5.22	0.29	5.46	3.74
χ^2 p-value		0.02	0.59	0.02	0.05
Panel C. Number of household members $= [7, 10]$					
Polluting fuel for cooking		-0.042	-0.004	-0.010	-0.027
		(0.029)	(0.007)	(0.017)	(0.024)
Speed of change in forest cover	0.023*				
	(0.013)				
Owns agricultural land	0.055***				
	(0.004)				
Observations	59,754	59,754	59,754	59,754	59,754
R^2	0.52	0.01	0.00	0.01	0.01
F-stat on IVs	90.37				
Hansen's J-statistic		0.00	0.13	2.65	1.75
χ^2 p-value		0.96	0.72	0.10	0.19

Notes: In this table, panel B combines the subsamples in panels B and C of Table S.1, while panel C combines the subsamples in panels D and E of Table S.1.

Figure S.2: Heterogeneous Effects of IAP on Infant Mortality by Child's Age and Household Size using IV (2SLS) Regressions for Households with Fewer than Twelve Members (3-by-3 Classification) (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level



Notes: The figure presents heterogeneous treatment effects of IAP (*defined by polluting fuel use*) on infant mortality by both child's age and household size using four distinct subpopulations of households with fewer than twelve members (generally based on our findings in Table 11) covered in NFHS-4 data. Panels (a)–(d) considers each of the four subsamples from 1-3 to 10-12 members in an orderly fashion. Each panel reports results from the estimation of Equation (1) using IV regression with different dependent variables and similar specifications where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The outcome variable in each IV regression is a binary variable of infant mortality for each of the four different age-groups, and the endogenous regressor (i.e., the outcome variable in the first-stage regression) is whether fuel choice: polluting fuel. The speed of change in forest cover and agricultural land ownership status are used as instruments, and first-stage coefficient estimates on the IVs are both positive and statistically significant at least at the 5% level except for the speed of change in forest cover in the first-stage regression for households with 10-12 members. The F-test on IVs verifies that the instruments generate a plausible variation in polluting fuel for cooking in all regressions. The Hansen's J-statistic suggests that the excluded IVs are not exogenous in under-five and post-neonatal mortality regressions for households with 4-6 members since a rejection of the null hypothesis of the Sargan-Hansen test is encountered at the 5% level. All specifications contain a vector of demographic controls and a 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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. Unit of observation: child. Heteroskedasticity-consistent standard errors are clustered by districts.

Table S.3: Heterogeneous Effects of IAP on Infant Mortality by Child's Age and Household Size using IV (2SLS) Regressions for Households with Fewer than Twelve Members (3-by-3 Classification) (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

	1st stage	2nd stage			
	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Panel A. Number of household members $= [1, 3]$					
Polluting fuel for cooking		0.196	0.020	0.089	0.088
		(0.146)	(0.038)	(0.083)	(0.121)
Speed of change in forest cover	0.046***				
	(0.015)				
Owns agricultural land	0.028***				
	(0.006)				
Observations	15,573	15,573	15,573	15,573	15,573
R^2	0.59	0.11	0.00	0.01	0.10
F-stat on IVs	14.73				
Hansen's J-statistic		1.99	0.73	1.27	1.66
χ^2 p-value		0.16	0.39	0.26	0.20
Panel B. Number of household members $= [4, 6]$					
Polluting fuel for cooking		0.150***	0.006	0.039***	0.105***
		(0.028)	(0.007)	(0.014)	(0.020)
Speed of change in forest cover	0.036***				
	(0.011)				
Owns agricultural land	0.051***				
	(0.003)				
Observations	104,818	104,818	104,818	104,818	104,818
R^2	0.55	-0.03	0.00	-0.00	-0.03
F-stat on IVs	112.43				
Hansen's J-statistic		4.17	0.01	4.39	2.25
χ^2 p-value		0.04	0.91	0.04	0.13
Panel C. Number of household members $= [7, 9]$					
Polluting fuel for cooking		-0.034	-0.001	-0.005	-0.028
		(0.029)	(0.007)	(0.018)	(0.025)
Speed of change in forest cover	0.025**				
	(0.012)				
Owns agricultural land	0.055***				
	(0.004)				
Observations	51,537	51,537	51,537	51,537	51,537
R^2	0.52	0.01	0.00	0.01	0.01
F-stat on IVs	85.75	13.07	2.14	756.46	18.29
Hansen's J-statistic		0.04	0.14	2.86	1.14
χ^2 p-value		0.85	0.71	0.09	0.29

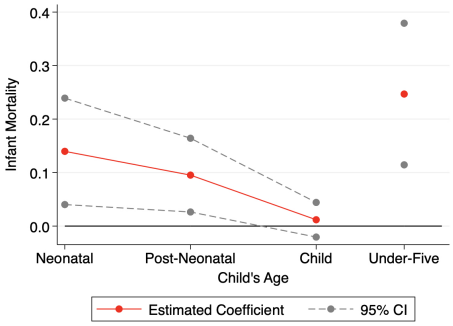
Table S.3: (Continued)

	1st stage	2nd stage			
	(1) Polluting Fuel Use	(2) Under-Five	(3) Child	(4) Post-Neonatal	(5) Neonatal
Panel D. Number of household members $= [10, 12]$					
Polluting fuel for cooking		0.015	-0.008	-0.018	0.041
		(0.048)	(0.013)	(0.024)	(0.039)
Speed of change in forest cover	0.000				
	(0.020)				
Owns agricultural land	0.071***				
	(0.008)				
Observations	16,927	16,927	16,927	16,927	16,927
R^2	0.52	0.01	-0.00	0.00	0.00
F-stat on IVs	35.77				
Hansen's J-statistic		0.37	0.28	0.00	0.29
χ^2 p-value		0.54	0.60	0.98	0.59

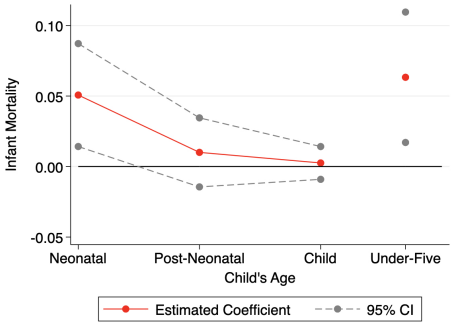
Notes: The table presents heterogeneous treatment effects of IAP on infant mortality by both child's age and household size using four subpopulations of households with fewer than twelve members (generally based on our findings in Table 11) covered in NFHS-4 data. Panel A–D considers each of the four subsamples from 1-3 to 10-12 members in an orderly fashion. The first column provides results from the first-stage regressions of the IV (2SLS) regressions, where the dependent variable is a binary variable of whether fuel choice: polluting fuel. Columns (2), (3), (4) and (5) report results from the estimation of Equation (1) using IV regression with different dependent variables and similar specifications where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The outcome variable in an IV regression is a binary variable of infant mortality for each of the four different age-groups. All specifications contain an unreported vector of demographic controls and a 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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. Unit of observation: child. Heteroskedasticity-consistent standard errors clustered by districts are in parentheses. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Figure S.3: Heterogeneous Effects of IAP on Infant Mortality by Child's Age and Household Size using IV (2SLS) Regressions for Households with Fewer than Twelve Members (4-by-4 Classification) (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

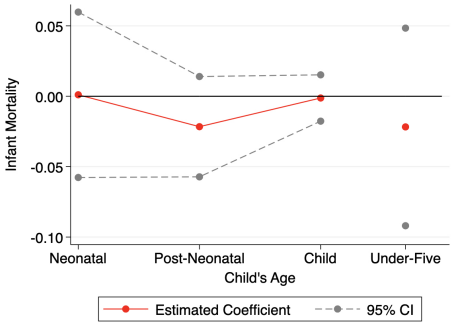
(a) Household size = $[1, 4]$, N = 49,401



(b) Household size = $[5, 8]$, N = 110,935



(c) Household size = $[9, 12]$, N = 28,519



Notes: The figure presents heterogeneous treatment effects of IAP (*defined by polluting fuel use*) on infant mortality by both child's age and household size using three distinct subpopulations of households with fewer than twelve members (generally based on our findings in Table 11) covered in NFHS-4 data. Panels (a)–(c) considers each of the three subsamples of 1-4, 5-8, and 9-12 members in an orderly fashion. Each panel reports results from the estimation of Equation (1) using IV regression with different dependent variables and similar specifications where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The outcome variable in each IV regression is a binary variable of infant mortality for each of the four different age-groups, and the endogenous regressor (i.e., the outcome variable in the first-stage regression) is whether fuel choice: polluting fuel. The speed of change in forest cover and agricultural land ownership status are used as instruments, and first-stage coefficient estimates on the IVs are both positive and statistically significant at the 1% level except for the speed of change in forest cover in the first-stage regression for households with 9-12 members. The F-test on IVs verifies that the instruments generate a plausible variation in polluting fuel for cooking in all regressions. The Hansen's J-statistic suggests that the excluded IVs are not exogenous in under-five and post-neonatal mortality regressions for households with 1-4 members since a rejection of the null hypothesis of the Sargan-Hansen test is encountered at the 10% and 5% level, respectively. All specifications contain a vector of demographic controls and a 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 educational attainment; and infant characteristics: gender and breastfeeding status. The state, year, and state-by-year dummies are also controlled in every specification. Unit of observation: child. Heteroskedasticity-consistent standard errors are clustered by districts.

Table S.4: Heterogeneous Effects of IAP on Infant Mortality by Child's Age and Household Size using IV (2SLS) Regressions for Households with Fewer than Twelve Members (4-by-4 Classification) (IVs = Speed of Change in Forest Cover and Agricultural Land Ownership): SEs clustered at district level

	1st stage	2nd stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting Fuel Use	Under-Five	Child	Post-Neonatal	Neonatal
Panel A. Number of household members $= [1, 4]$					
Polluting fuel for cooking		0.247***	0.012	0.095***	0.140***
		(0.068)	(0.017)	(0.035)	(0.051)
Speed of change in forest cover	0.039***				
	(0.012)				
Owns agricultural land	0.036***				
	(0.004)				
Observations	49,401	49,401	49,401	49,401	49,401
R^2	0.58	-0.01	0.00	-0.02	0.01
F-stat on IVs	40.24				
Hansen's J-statistic		3.57	0.95	4.93	2.10
χ^2 p-value		0.06	0.33	0.03	0.15
Panel B. Number of household members $= [5, 8]$					
Polluting fuel for cooking		0.063***	0.003	0.010	0.051***
		(0.024)	(0.006)	(0.012)	(0.019)
Speed of change in forest cover	0.032***				
	(0.011)				
Owns agricultural land	0.053***				
	(0.003)				
Observations	110,935	110,935	110,935	110,935	110,935
R^2	0.53	0.01	0.00	0.01	0.00
F-stat on IVs	132.53				
Hansen's J-statistic		1.98	0.00	0.31	2.61
χ^2 p-value		0.16	0.99	0.58	0.11
Panel C. Number of household members $= [9, 12]$					
Polluting fuel for cooking		-0.022	-0.001	-0.022	0.001
		(0.036)	(0.008)	(0.018)	(0.030)
Speed of change in forest cover	0.008				
	(0.016)				
Owns agricultural land	0.069***				
	(0.007)				
Observations	28,519	28,519	28,519	28,519	28,519
R^2	0.53	0.02	0.00	0.00	0.01
F-stat on IVs	55.30				
Hansen's J-statistic		0.01	0.27	1.01	0.41
χ^2 p-value		0.92	0.60	0.31	0.52

Notes: The table presents heterogeneous treatment effects of IAP on infant mortality by both child's age and household size using three subpopulations of households with fewer than twelve members (generally based on our findings in Table 11) covered in NFHS-4 data. Panel A–C considers each of the three subsamples from 1-4 to 9-12 members in an orderly fashion.